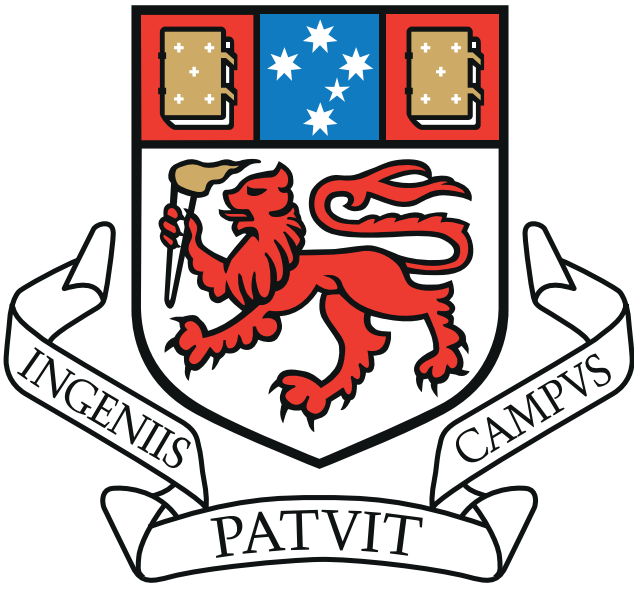
Dig a little deeper and it gets weirder: Social Media and Manipulation

Jonathon Alan Manning, BComp Hons

A dissertation submitted to the
School of Computing and Information Systems
in partial fulfilment for the degree of
Doctor of Philosophy



University of Tasmania

November 4, 2015

Declarations

Originality

I hereby declare that to the best of my knowledge, this thesis has not been submitted for the award of any diploma or degree at any other tertiary institution. It is also my belief that the thesis contains no previously published material except where due reference is made.

Jonathon Alan Manning
November 4, 2015

Access

This thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Jonathon Alan Manning
November 4, 2015

Abstract

The research presented in this thesis provides a detailed discussion and analysis of the phenomenon of *content ranking manipulation* on social media sites, in which the ranking of content submitted to social media sites is artificially manipulated in order to increase its prominence, at the expense of other content submitted to the site. The thesis documents the various types of manipulation that were identified on Reddit, a popular social media site; in addition, the impact of this manipulation is discussed in the context of different types of communities that exist inside social media sites. A framework for discussing manipulation and its impacts upon social media site users is proposed, and is discussed in the context of existing social media sites.

Over the last decade, social media sites have rapidly risen to prominence as one of the most popular types of site on the World Wide Web. Sites such as Facebook, Twitter and Reddit act as a hub in which users may access and read "content" - for example, links to articles, photographs, videos and more - while at the same time submitting new content to the site. Social media sites are largely un-curated: users do not have to be approved by the administrators before they are able to add content to the site. In order to provide some means of ensuring that submitted content is high-quality, social media sites typically use some form of quantitative measurement to determine how prominent each piece of submitted content should be on the site. This is typically implemented using a voting mechanism, in which users are allowed to vote on individual pieces of content, and the votes are aggregated to determine the ranking. On social media sites, users who are submitting content to the social media site also have the ability to vote on content submitted by others. Due to the de-

sire of submitters to have their content be prominently displayed, various techniques may be employed in order to ensure that *their* content receives more votes than *other's* content. These techniques constitute the types of manipulation identified in this research and presented in this thesis.

The existing literature features several examples of studies into manipulation; however, these studies focus on in-depth analysis of specific manipulation techniques, such as Douceur's (2002) discussion of the Sybil attack. By contrast, this research takes a significantly broader approach to the discussion of manipulation by first identifying a definition of what manipulation *is*, based on interviews with community administrators and moderators, and then applying social research techniques to determine what kinds of manipulation exist. As a result, this thesis identifies new kinds of manipulation and enables further, more focused research on manipulation.

The research takes the form of a three-phase study, in which administrators, moderators and users of a large social media site participated. Methods based on a grounded approach to qualitative data analysis were extensively employed in this research, and were used in the analysis of the data collected in all three phases. In the first phase of the research, administrators and moderators of Reddit were interviewed, and analysis revealed a number of high-level categories of manipulation seen by these users. The second phase of the research employed an innovative data collection tool that operated inside the web browser of all participants in the study, which gathered additional information on end-user perspectives of manipulation; this served to both confirms the completeness of the categories of manipulation identified in the first study, and to provide more detail for each category of manipulation using user perspectives of manipulation. The third phase of the research involved interviewing partici-

pants from the second phase, and gathered data regarding the impact that manipulation had on their experience of the social media site.

Analysis of the collected data was performed using a social research approach, based on grounded-theory based methods to derive a systematic analysis of the data gathered in all three phases of the research. The resulting analysis is then discussed, and findings are derived.

The research has identified a broad variety of different forms of social media site manipulations, each of which is broken down into multiple sub-categories and discussed. Each of these categories has a variety of different forms, which are discussed in this thesis, and examples presented. The research found that users generally share the same negative views of manipulation as administrators and moderators do, but note that some kinds of manipulation actually has a positive impact on their experience of the site.

The research relates manipulation to earlier work done on the subject of *relevance* (Saracevic 1975, 2007), and establishes that attempts to manipulate the ranking of a social media site are attempts to modify the systemic relevance of content for all users by modifying other forms of relevance (such as cognitive, affective, and topical) for smaller groups of users.

The contributions of this research are an identification of categories of manipulation on social media sites, a discussion of the design, implementation and results of a novel approach to data collection in an online study, and a model of manipulation and its impact upon users of different kinds of online communities. These contributions extend and broaden the scope for future research into social media site manipulation as a wide topic, which has previously been limited to individual forms of manipulation.

Acknowledgements

First and foremost: thank you to every single participant who took part in this research. This thesis couldn't have happened with you.

Special thanks to Max Goodman, who helped me arrange access to Reddit. Thanks also to the rest of the team at Reddit: you guys do some amazing work.

Massive, massive thanks to my two supervisors: Professor Christopher Lueg, and Dr Leonie Ellis. You've both been amazingly helpful, supportive, patient, and insightful.

Thanks to Tony Gray, and the AUC (http://www.auc.edu.au), which is one of the biggest reasons I am where I am right now.

Thanks to the peanut gallery that is the folks of "Maclab", for keeping me sane throughout all of this. Congratulations, you guys! We made it. Now we're all unemployable.

Huge, huge thanks to Dr Paris Buttfield-Addison, my friend and business partner at Secret Lab (http://secretlab.com.au) for both the tremendous support and advice in writing this thesis, and in keeping the business running while I was working on it.

Thanks also to the multiple places where I've had an opportunity to work on this thesis: Hobart, Melbourne, Sydney, Brisbane, Auckland, Hong Kong, San Francisco, Mountain View, Palo Alto, Portland, Seattle, New York City, Chicago, Denver, Boulder, London, Austin, Los Angeles, Las Vegas, New Orleans, Montreal, Malmö, and Copenhagen, amongst others.

Very special thanks to Ash Johnson. You're great.

Finally, lots of love to my very large, very extended family. Told you I'd get it done.

Typeset in LaTeX.

Contents

List of Figures

1 Introduction

1.1 What's the big deal with social media, anyway?

Social media sites are some of the most popular sites on the internet. The most famous of these, at the time of writing, is Facebook, with 1.15 billion users Facebook Inc. (2013). Social media sites are so named due to their *social* nature, in that they are intended to facilitate the sharing of media between people (Agichtein, Castillo, Donato, Gionis and Mishne 2008). In this thesis, the term 'social media site' refers to sites in which the content of the site is determined by the users of that site; in cases typified by such sites as Reddit (Reddit Inc. 2014b) and MetaFilter (Metafilter Network Inc. 2013a), the *only* content shown by the site is that which the audience of the submits.

Because the content that appears on these kinds of sites are controlled by the users, and because there is a limited amount of space on the site (both in terms of screen real estate and the amount of time a user will want to spend browsing the site), competition exists between users who

want *their* content to appear.

Many social media sites rank submitted content based on voting; examples include Reddit (Reddit Inc. 2014b) and Hacker News (Y Combinator Inc. 2014). For each piece of content submitted to the site, users vote up or down. These votes are then used by the software running the website to determine what content is shown on the front page of the site. The algorithms used vary from site to site – for example, Reddit uses an algorithm that combines the net total of upvotes minus downvotes, weighted based on the age of the content (Salihefendic 2010).

Because the users of social media sites determine the content of the site, social media sites are subject to manipulation and abuse. What is unclear, however, was what specific behaviour administrators and moderators of social media sites consider to be manipulation. Most social media sites have a "code of conduct": Reddit has its "*reddiquette*" , while MetaFilter has a general FAQ that covers, among other things, posting etiquette (Metafilter Network Inc. 2013b). These represent codified rules of conduct for communities, and provide specific instructions for users on what to do and what not to do.

1.1.1 Past research

Social media sites have been the subject of academic research for a long time; studies have ranged from interpersonal dynamics in social media sites (Hogg and Lerman 2012), to the effects of ephemerality and anonymity (Bernstein, Monroy-Hernández, Harry, André, Panovich and Vargas 2011), to attempts to detect and measure contribution quality (Diakopoulos and Naaman 2011). Work has also been done in the area of manipulation in social media sites; however, most work has focused on specific attacks, such as Douceur's (2002) Sybil attack and John, Yu, Xie, Krishnamurthy and

Abadi's (2011) work on search-engine result poisoning. However, very little qualitative research has been done into the wider-scale phenomenon of manipulation on social media sites. This work extends the quantitative work already present, and provides a context for understanding manipulation in social media in general, by gathering experiences from both site administrators and users on how they perceive and define manipulation.

1.1.2 A note regarding coarse language

This thesis presents several quotations taken from social media sites, and focuses on a topic with which several participants expressed strong frustration. These expressions often include swearing or other coarse language. In order to preserve all relevant sentiments that were expressed, profanity and offensive language have not been removed.

1.2 Research Questions

This study began with the following questions in mind. These questions are frequently referred to in later chapters, and as such have been given identifiers as a shorthand.

RQ1 What are the most prevalent kinds of manipulations taking place on these social media sites?

RQ2 What impact do these types of manipulations have on the communities?

RQ3 How severe are the different types of manipulations in terms of their impact on a community?

RQ4 What can site owners do to address manipulation?

In this thesis, answers to the above questions are sought through *a*) interviews with social media site administrators, to gather administrator perspectives (presented in Chapter 3); *b*) a diary study conducted among social media site users, to gather user perspectives (presented in Chapter 4); and *c*) interviews with social media site users, to provide additional context to the information gathered in the diary study (presented in Chapter 5).

Each of these chapters incrementally builds an understanding of manipulation in social media sites; findings from each of the chapters are then synthesised, and implications for site administrators and researchers in this field are are presented in Chapter 6. This structure allows the reader to follow the ideas presented in Chapter 6 as they develop.

1.3 Contributions

This research extends the existing literature by providing the following contributions:

- an understanding of user and administrator perceptions of manipulation;
- an identification and classification of manipulation methods;
- an understanding of the impact of manipulation methods, and their severity; and
- a set of best practices for tool developers aiming to reduce the negative impact of manipulation.

These contributions are discussed in detail in Chapter 6.

1.4 Thesis Structure

This remainder of this thesis is presented according to the following structure:

- Chapter 2 presents the literature review.
- Chapter 3, Chapter 4 and Chapter 5 each present the conduct and outcomes of the three studies into manipulation in social media sites conducted during this research.
- Chapter 6 discusses and synthesizes the outcomes of these studies, and a framework for discussing manipulation in social media sites is presented.
- Chapter 7 presents the contributions of this thesis, and concludes with parting words.

2

Literature Review

"An SEO expert walks into a bar, bars, beer garden, hangout, lounge, night club, mini bar, bar stool, tavern, pub, beer, wine, whiskey... (Moynihan 2012)"

This chapter presents a critical review of the context in which this research is conducted, which provides the reader with an understanding of the most important aspects in this complex field. This chapter highlights exactly what is being researched, why it is being researched, and what is currently missing in the literature. In addition, the methodology of past research in this area is presented and discussed.

2.1 Introduction

In order to discuss manipulation in social media sites, it is important to have a solid background in certain key topics. This literature review begins by providing background information on several social media sites that are repeatedly cited as examples in this thesis, and then proceeds to explore past research that relates to these sites and sites like them.

This chapter takes the following structure:

- Section 2.2 dicusses the social media websites that serve as examples in this thesis. In particular, Reddit, the site that formed the case study in this research, is introduced.
- Section 2.3 introduces social media, and discusses several important topics related to it: the concept of content quality, wear, social interaction history and social navigation, information behaviours, exploratory searching, and the theory of relevance.
- Section 2.4 discusses the ranking systems used by the websites discussed in Section 2.2, and how they differ among each other.
- Section 2.5 discusses recent work into the specific topic of manipulation of social media communities.

2.2 Social media communities discussed in this thesis

The research presented in this thesis focuses on the behaviour of users in social media communities. In particular, a specific social media community, *Reddit*, was the case study for the research. In order to provide context for the sites discussed in this document, this section presents and discusses the various social media sites that are referred to. Alongside these discussions of each of the sites, the evolution of the aspects of popular social media communities is discussed.

2.2.1 Usenet

One of the best studied online communities is *Usenet*, a distributed forum system that rose to prominence in the early 1980s (Lueg and Fisher 2003). Usenet is a popular topic for research into online communities and social networking, owing to its age, openness, and volume of activity (Wellman, Salaff, Dimitrova, Garton, Gulia and Haythornthwaite 1996, Harrison and Dourish 1996, Fisher, Smith and Welser 2006, Bury, Deller, Greenwood and Jones 2013).

Usenet is a decentralised bulletin-board messaging system. Unlike the other examples in this section, which rely on a central server, messages may be posted from a Usenet client (such as Mozilla Thunderbird, seen in Figure 2.1) to any Usenet server; these are then relayed across the Usenet network to other servers, and eventually retrieved by end users (Lueg and Fisher 2003).

Usenet serves as an excellent starting point from which to begin a discussion of the manipulation of social media sites. While the decentralised nature of Usenet differs from the generally-accepted definition of "social media sites", Usenet shares many important traits with the communities that this thesis is primarily concerned with. While it was not the very first online social network, the popularity of Usenet in the 1980's and 1990's caused to to serve as the prototype for many later systems, such as Tumblr (Bury et al. 2013).

Usenet is a *thread-based* discussion system. Users create a new thread of discussion by posting a message to a Usenet server. This message is then distributed across the Usenet network, which allows users to download that message and view it. What makes Usenet a thread-based discussion system is that the system allows for users to post messages that are replies

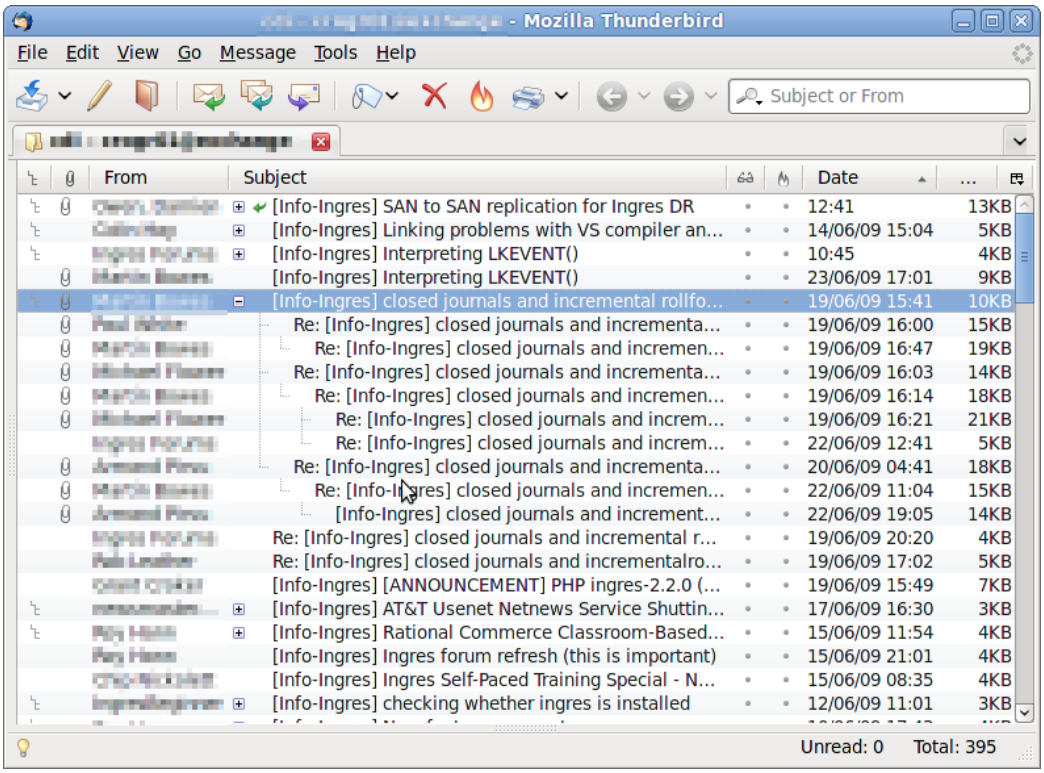


Figure 2.1: Mozilla Thunderbird, a Usenet client.

to previous messages. Users can then reply to another's reply, and so on.

Having a thread-based discussion system means that an online community does not have to restrict itself to a single conversation at a time; by breaking up the ongoing conversations into multiple areas, the community can discuss many different topics at the same time. This means that the population of the community is able to grow, because it is more likely for a visitor to see a topic that they wish to discuss.

2.2.2 Web forums

A *web forum*, often shortened to simply *forum*, is an environment in which users may post new discussion threads, and replies to that thread. Web forums are not a single site, but rather a category of site. Most web forums

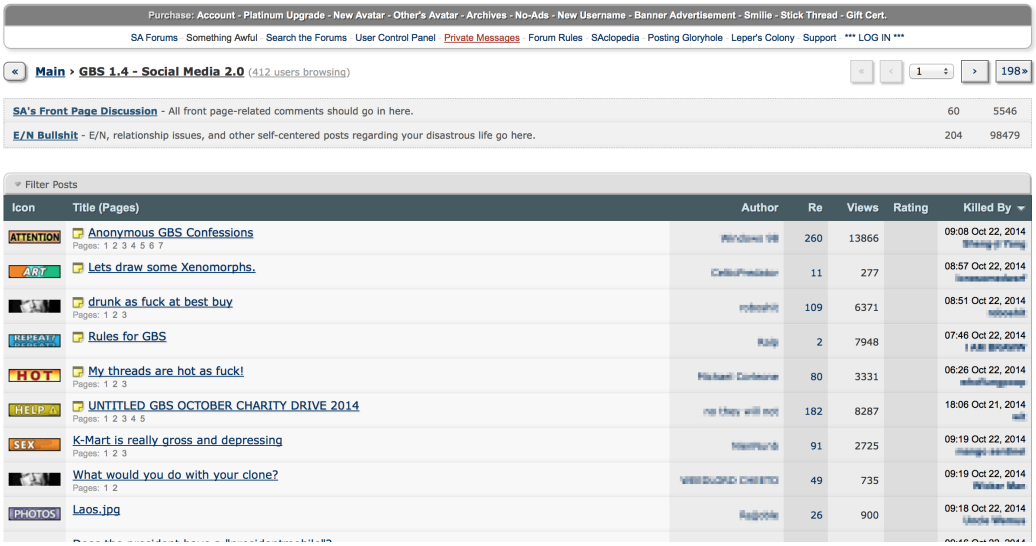


Figure 2.2: Something Awful, a web forum.

are sites that run one of several different software packages, popular examples of which include phpBB (phpBB Team 2013), vBulletin (vBulletin Solutions 2014) and Invision PowerBoard (IPS Inc. 2013). An example of a web forum site is Something Awful (Something Awful LLC 2014b, shown in Figure 2.2), which uses a customised version of the vBulletin software.

A web forum is distinguished by the following elements in its user interactions. Users may post new discussion threads, or replies to existing discussion threads. The prominence of a discussion thread depends on the time of the most recent reply. Replies to a comment thread are displayed in chronological order.

Individual web forums are structurally very similar to Usenet, in that they are arranged around threads to which users post additional comments. Their most striking difference is that while Usenet is a distributed system, web forums are centralised, and each forum is under the control of relatively few individuals. This means that the administrators of these websites have significant control over the nature and content of the com-

munity.

2.2.3 Slashdot

Slashdot (Dice Holdings Inc. 2014, shown in Figure 2.3), is a social media site that styles itself around "news for nerds". The format of the site involves news stories being posted by site staff; each story is available to be commented on by users. Slashdot's moderation system is complex, featuring multiple different types of moderation, allocated moderation points, and meta-moderation.

Users are allowed to submit articles (called "posts" on Slashdot) to the site owners for consideration, but the selection of which posts are displayed on the site and which are rejected is controlled by humans, and not by an algorithm. The process of submitting comments that are attached to stories, however, is not human-mediated, and any user may submit a comment, either anonymously or attached to their account on the site. If a user posts a comment anonymously, their username is given as "Anonymous Coward". The slightly pejorative phrasing for this anonymous moniker is deliberate, as it implies that, by posting anonymously, the user is not taking full responsibility for their comments. (Smith 2011)

Slashdot has been studied extensively as an early example of user-controlled content moderation on a public website. As a centrally controlled website, Slashdot is able to make changes to its content moderation system that immediately apply to all users, whereas the distributed nature of Usenet makes such wide-scale changes impossible. Along these lines, Lampe, Johnston and Resnick (2007) note that while Slashdot permits its users to apply their own personal preferences for filtering content based on ranking, few take advantage of this; they suggest that operators of similar websites should take note of the preferences of users who express filter

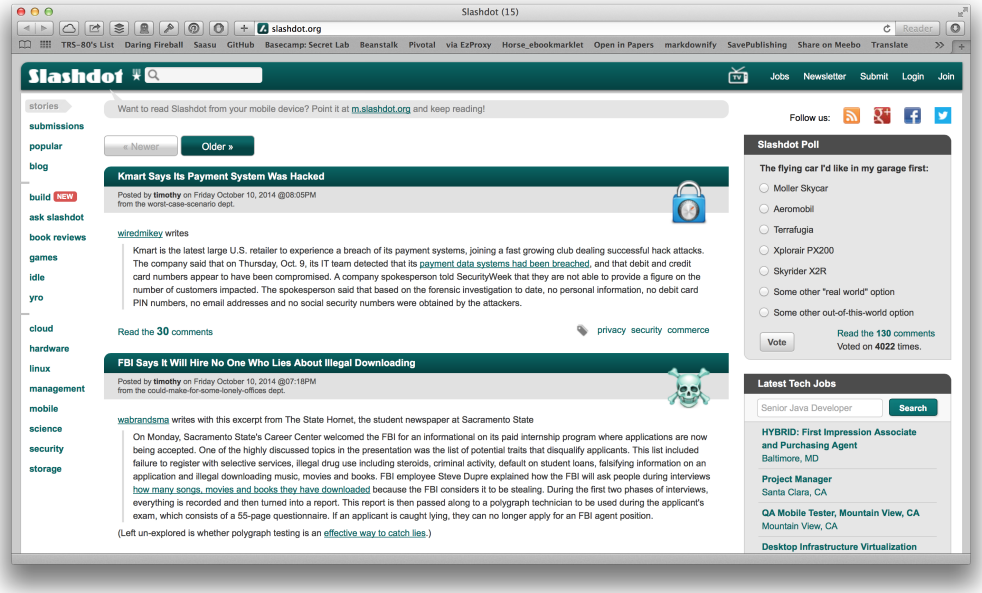


Figure 2.3: Slashdot.

preferences, and consider allowing these preferences to influence the default settings for all users.

Because so much of Slashdot's content is heavily annotated and easily quantified, it becomes possible to make predictions on the popularity of content. Kaltenbrunner, Gomez and Lopez (2007) were able to not only predict the number of comments applied to content, but also the approximate times at which such comments were likely to appear.

Slashdot is an interesting example of a web forum-like community that allows users partial control over the site's content. Adapting terminology from web forums, we may say that users have no control over the threads, but significant control over the replies.

2.2.4 Digg

Digg (shown in Figure 2.4), is a social news site founded in 2004 (News.me Inc. 2014) that catered towards a similar audience as Slashdot. Digg was originally a site in which users could post links to content on the internet, comment on it, and vote on that content. In 2012, following an acquisition of the company that created it, Digg was re-launched as an editorially driven news aggregator, in which users do not have direct control over the site.

In its former incarnation, Digg was a popular topic for researchers in the field of social media and user-controlled websites. Digg has been the subject of considerable past research (Wu and Huberman 2007, Zhu 2009, Szabo and Huberman 2010, Hogg and Lerman 2012); however, the 2012 relaunch rendered much of the Digg-specific elements of previous research no longer applicable. Conclusions made in previous work that studied Digg still apply to the broader discussion of social media sites, but Digg-specific discussion from past papers is now no longer applicable to the site as it currently exists.

Digg presented another interesting aspect of the evolution of social media sites: in addition to allowing users to vote on and moderate comments on the stories that were posted to the site, Digg also gave users the same kind of control over the stories themselves. This removed the need to have a dedicated editorial team, as seen on Slashdot, and gave the community the ability to choose its own editorial direction.

2.2.5 Reddit

Reddit (shown in Figure 2.5) is a community-oriented social networking and social media site founded in 2005 (Reddit Inc. 2014c), in which users

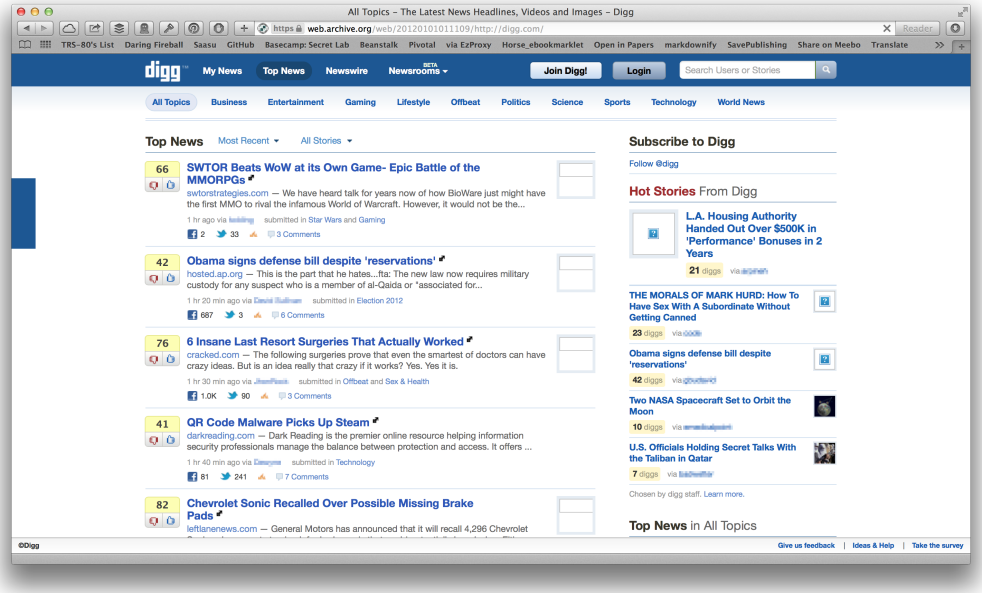


Figure 2.4: Digg, in its pre-2012 incarnation. A static archive of the site is preserved by the Internet Archive (2012).

may submit content, or *posts*, such as text posts or links to other locations on the web, and reply to that content in the form of comments. Users may also vote on posts (as seen in Figure 2.6) and comments; these votes are either *upvotes* or *downvotes*. When users vote on posts or comments, they become more or less prominent.

Because Reddit relies on user votes to determine what content is considered to have sufficiently high quality to appear on the site, a lack of voting by users who rely on others to do the voting can result in potentially popular content being ignored (Gilbert 2013).

Reddit also allows users to create *sub-reddits*: sub-sections of the site, into which conversation threads are grouped. For example, reddit contains separate subreddits for discussing atheism, science, for posting amusing pictures, and for discussing world news. The ability to create subreddits is not restricted to site staff, but rather is available to all registered

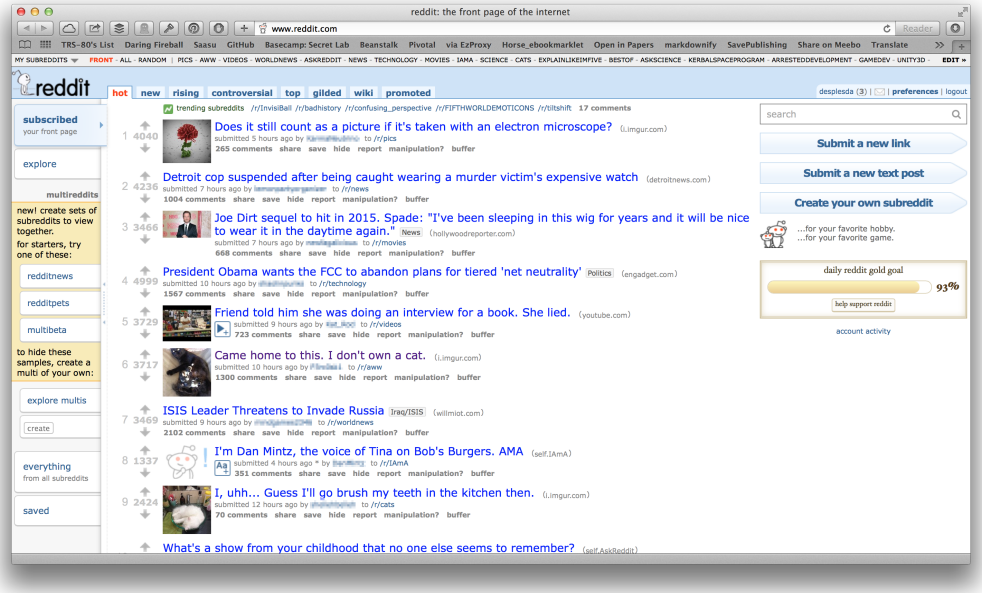


Figure 2.5: Reddit.

Figure 2.6: A post on Reddit. The voting controls and total net score are at the left; in this image, the user has selected the upvote arrow, increasing the score of the post by one vote.

users of the site.

Subreddits are may be viewed in isolation, but users may choose to view multiple subreddits at the same time. For example, the front page of reddit.com is actually the combination of the top 20 subreddits, selected by the site administrators.

This has had the effect of turning Reddit into a "platform" for communities: if a group of like-minded people wish to come together to discuss a topic, a single person may create a subreddit for that topic.

Reddit, as is the focus of this research, presents another interesting evo-

lution from Digg's model: in addition to giving users control over the site's content, it allows users to create their own separate sub-communities. This permits users who are especially interested in a topic to post as much about it as they please, without risking annoying other users; in addition, this ability to create separate sub-communities allows users to present and discuss their topic without this discussion being lost in the larger stream.

2.3 Background

In order to discuss social media manipulation, one must first have an understanding of the development and background of the wider topic of social media sites. This section provides this background by presenting past research into online communities, and the concept of user-generated content, against which this research is situated. The literature relating both practical and theoretical work in the areas of social media is discussed in detail, along with relevant adjacent fields.

2.3.1 Online communities

Some of the earliest research into online communities began with analysis into the behaviour of office workers who began to use networked computers to augment paper-based information management. Examples of this research include Hiltz (1985), who discussed various prototypical systems designed to improve worker productivity in office environments.

Computers are used by people; computers are also networked together. As Wellman et al. (1996) notes, computer networks become *social* networks when they connect people to other people, as well as to computers. Computer-connected social networks enable the formation of online communities; early work done by Rheingold (1993) is one of the first efforts to describe

the sociological phenomenona of "virtual communities" as communities of individuals that interact through online networks.

Much of the early research into virtual communities explored how collections of indivudals developed communtities with their own distinct cultures, norms, and senses of social connection. For example, Baym (2000) discusses the development of shared norms that developed among a Usenet newsgroup that focused on soap operas; Rheingold (1993) discusses *The WELL*, an early virtual community that acted as the prototype for many current social media sites.

The underlying definition of "community" is difficult to pin down. Indeed, as Komito (1998) notes, the term "means many things to many people, and it would be hard to find a definition of community that would be widely accepted". Indeed, Hillery (1955) found over 94 different definitions of the term. This thesis uses the definition of *virtual community* proposed by Porter (2006): "an aggregation of individuals or business partners who interact around a shared interest, where the interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms."

Online communities are a type of 'place' that does not involve a spatial component. Harrison and Dourish (1996) note that "a virtual space only presents only the opportunity for a virtual place to develop". Harrison and Dourish go on to mention that Usenet, discussed in Section 2.2, serves as an example of a "placeful" area that lacks physical space.

The content of social media communities is often in a state of rapid flux, with the content of some sites being almost ephemeral. Certain online communities, such as 4chan (4chan, LLC 2014), an anonymous discussion forum with high volumes of user activity, make use of *thread expiration* (Bernstein et al. 2011), in which new threads appear at the top of the list,

and are pushed down as new threads are added. The number of threads per page is limited, meaning that older threads move off the first page as new threads are added. If a user replies to a thread, it is bumped back to the top of the first page. Bernstein et al. notes that a thread can progress from the top of the first page to the bottom of the *fifthteenth* page on 4chan in under a minute.

2.3.2 Social media

The term "social media" covers a broad array of different types of sites and communities, though all are linked by the central theme of the users of the site generating the content of the site, in addition to those users interacting with each other within the context of the same site (Kaplan and Haenlein 2010). In this sense, social media sites act as both a repository of user-created content and as a platform that enables a community centered around that content.

In the context of this thesis, *content* is used as a generic term to refer to text, images, video, audio, links to websites, or any other information made available through a website (Vickery and Wunsch-Vincent 2007). Some content is *user-generated*, which is of particular interest when discussing social media.

The exchange of user-generated or user-created content is the primary purpose of many social media sites, such as Digg, YouTube, MySpace, Reddit and more (Vickery and Wunsch-Vincent 2007, Guo, Tan, Chen, Zhang and Zhao 2009, Zhu 2009, Szabo and Huberman 2010, Shneiderman, Preece and Pirolli 2011, Garcia, Mendez, Serdült and Schweitzer 2012). User-created content, occasionally referred to as user-generated content, is content created by the users of a community for the benefit of other users of the same community (Vickery and Wunsch-Vincent 2007).

The precise definition offered by Vickery and Wunsch-Vincent involves three characteristics, which are summarised here.

1. User-generated content is *published*. The content is available to other people via a website or some other online resource; this critierion does not preclude the content from only being available to a limited number of users, such as only being available to students at a particular university[1].

2. User-generated content involves a degree of *creative effort*. A photo found on the internet and uploaded to a website without modification is not user-generated content; in order to qualify as user-generated, the user who uploads it must have contributed some creative effort to it.

3. Finally, user-generated content is *created outside of a professional context*. User-generated content is not created with the primary expectation of a commercial interest; rather, user-generated content is created and distributed with other, more personal goals in mind, such as connecting with one's peers, achieving fame or notoreity, or self-expression. Note that this criterion does not restrict user-generated content creators from being professionals; a professional photographer may share their photographs on a social media site; it only ceases being user-generated content under this definition when they charge money for access.

User-generated content is submitted to social media content in different ways by different users. Different patterns of contribution were identi-

[1] Vickery and Wunsch-Vincent note that, in theory, user-generated content *could* remain unpublished and never made available to anyone but its author; however, considering only content that is published allows the definition to exclude private content such as email.

fied and modelled by (Guo et al. 2009). On a related note, Anderson, Huttenlocher, Kleinberg and Leskovec (2012b) identified that the evaluation of other *users* in a social network site (as opposed to simply the content posted by these users) could be predicted by their relative status within the community.

Olson and Neal (2013) discusses mapping relationships between communities in Reddit, and demonstrates a method of visualising Reddit as a whole. As a result, community clusters can be identified, which allows for the identification of closely-related "meta-communities". The identification of these meta-communities opens up opportunities for meta-analysis of community content and behaviour, as well as providing additional data to determine the relative size and importance of individual communities.

In discussing this variety of social media sites, Kaplan and Haenlein (2010) created a classification of social media communities and varies across two axes: first, the degree of social presence (Short, Williams and Christie 1976) of the users, combined with the degree of media richness present on the site (Daft and Lengel 1986); secondly, the degree of self-presentation (Goffman 1959) and self-disclosure that a community requires in order to participate. While Goffman (1959), Short et al. (1976) and Daft and Lengel (1986) all pre-date the emergence of social media sites, their analysis of human behaviour is just as applicable to the online space.

2.3.3 Different types of online communities

The framework constructed by Kaplan and Haenlein allows for the classification of existing social media sites, as shown in Figure 2.7. In the context of this classification, Reddit may be classified as requiring low degree of self-disclosure, and a medium degree of media richness; it may therefore be classified among other content communities, such as YouTube,

SoundCloud (SoundCloud Ltd. 2014) and other content-centric social media sites.

Reddit requires *low self-disclosure*: users are not required to disclose any information about themselves in order to join, with the exception of selecting a short, often pseudonymous username. Using the terminology from Preece and Shneiderman (2009), in order to be a "contributor", one does not need to disclose any personal information whatsoever.

The degree of *media richness* afforded by Reddit is moderate: Reddit's content is capable of facilitating significantly more rapid interpretation of cues than that afforded by a blog, but provides less rapid social feedback to participants in a telephone call. The fact that all participants in the Reddit community are able to directly address each other in a rich and detailed way allows for the creation of a personal focus of the content (Lengel and Daft 1988). This has implications for both the impact of and the methods of manipulation; for example, the "reference to self" type of attention-grabbing manipulation, which will be discussed in detail in Section 4.4.1. requires the ability for content to have the ability to clearly focus on the identity of individual participants, and to encourage the reader to feel empathy towards the subject.

Users have different roles in a social media community. Preece and Shneiderman (2009) discusses the "reader-to-leader" framework of user participantion, which presents the progression of users from passive consumers through contribution and finally to a leadership position within the social media community. In presenting the reader-to-leader framework, Preece and Shneiderman discusses the conditions under which a users will develop from one phase of their membership (for example, in deciding to become a reader, or in transitioning from a passive reader to a more active contributor). This "reader-to-leader" terminology is echoed

Self Presentation / Self Disclosure	Social Presence / Media Richness: Low	Medium	High
High	Blogs	Social network sites *(e.g., Facebook)*	Virtual social worlds *(e.g., Second Life)*
Low	Collaborative projects *(e.g., Wikipedia)*	Content communities *(e.g., YouTube)*	Virtual game worlds *(e.g., World of Warcraft)*

Figure 2.7: The classification of social media sites, from Kaplan and Haenlein (2010). Reddit, the focus of this thesis, is a *content community*, in which users share links, videos, photos and text. It requires low self-disclosure, and has a medium level of social presence and media richness.

by Kumar, Novak and Tomkins (2006), in which the analysis of social networks present in social media sites revealed users who were entirely passive members in the network, users who invited their friends to join the network, and users who fully participated in the network.

Social media sites run the risk of becoming internally fragmented into sub-communities (Van Alstyne and Brynjolfsson 2005). Users may retreat into sub-groups of the wider community that reflect their own worldview, and which may limit interaction with other viewpoints. This "balkanization" of social media sites, to use Van Alstyne and Brynjolfsson's term, is not avoided but embraced in the design of Reddit, which allows users to create their own sub-forums (*"subreddits"*) that focus on topics of their choosing; at the time of writing, there were over 25,000 user-created subreddits (Subreddits.org 2014), with more being created every day.

Gilbert (2013) found that 52% of the most popular content submitted to Reddit did not receive upvotes the first few times they were posted, and only received sufficient votes to make it to the front page. In their follow-up discussion of why this was the case, Gilbert speculated that ranking manipulation techniques may be in play; however, in this case, Gilbert's

speculation was that they were not appearing because of a *lack* of the use of techniques identified in this thesis as manipulation.

Konstan and Chen (2007) discusses field experiment design methodologies in social media contexts, and provides advice for researchers looking to conduct research in online social media sites. Various different modes of data collection are discussed, including deliberately posing as a normal user, scraping data after a particularly interesting event, creating a site specifically for the purposes of research, and collaborating with site owners. Konstan and Chen provided the inspiration for the researcher's decision to reach out directly to site administrators and moderators on Reddit in order to gather the data presented in Chapter 3 and Chapter 4.

2.3.4 Content quality in social media sites

Content that is submitted to social media sites varies in quality. This concept, while intuitive, is a foundational problem explored by the research presented this thesis, and consequently deserves elaboration.

The question of what defines 'quality' depends upon the site; Agichtein et al. (2008) provide a discussion of *objective* quality in the context of Yahoo! Answers (Yahoo! Inc 2005), and presents a framework for classifying the quality of content submitted to social media sites. Agichtein et al. note, however, that their study was limited to question-and-answer sites, in which quality is easier to define (for example, in terms of grammatical correctness, completeness of information, and so on) than in general content communities (as defined by Kaplan and Haenlein in 2010).

In an ideal situation, the most prominent content on a social media site would always be the "best", or highest-quality, content. However, as Agichtein et al. (2008) note, content submitted to social media sites vary significantly in quality, from high-quality content to "low-quality, some-

times abusive" content.

Content quality has been explored in the past: two examples of social media sites that have seen past research are Wikipedia, an online encyclopedia project that allows all users to make modifications to articles, and online question-and-answer (Q&A) sites: Kittur and Kraut (2008) found that implicit coordination in editing Wikipedia articles was more useful in improving article quality than explicit planning of coordination, while Harper, Raban, Rafaeli and Konstan (2008) found that reducing the barriers to entry for potential contributors increased overall answer quality on Q&A sites.

2.3.5 Wear

When physical objects are used, evidence of this use is left in the form of physical wear marks: metal is scraped, paint is chipped, and other mild, unavoidable damage mars the formerly pristine surface of the object.

This concept of *wear* is applicable to information systems; the concept as applied to information, as elucidated by Hill, Hollan, Wroblewski and McCandless (1992), is as follows: as users interact with information, they leave behind traces of those interactions. These traces are referred to by Hill et al. as wear, which describes two kinds of wear: *read wear*, and *edit wear*.

Read wear is recorded information that indicates patterns of access to information by users. This information is then able to provide later users of the system with additional context and history; the example given by Hill et al. is the use of an augmented scrollbar attached to a document, which records and displays where past users have scrolled to, and thereby provides an indication of which parts of the document are the most heavily read. Read wear is of particular relevance when one seeks to understand

patterns of user interaction (Hill et al. 1992).

By contrast, *edit wear* is recorded information that indicates the history of how other users have modified information. Kaptelinin (2003) serves as an excellent example of edit wear in action, in which is discussed a system that derives and displays the history and context of a project based on the editing patterns of the users.

Most kinds of wear are passive, and no particular action is required to be taken by the user in order for wear to be created (Hill et al. 1992). This allows tools that make use of wear to be useful without imposing additional burden upon the user.

Social media sites offer a rich field of possibilities for the collection of edit wear. To take Reddit as an example:

- Users create *posts*, which contain links to content and text; posts contain links to their creator's account, the time and date of their creation, and are situated in a specific section of Reddit (a "subreddit"). This additional metadata allows for the discussion of trends of content, and analysis of which areas of interest on the site are the most active. This information is used by Reddit itself in creating recommendations for new subreddits that users may be interested in; third-party analyses of this information also exist, such as Metareddit (2014).

- Users can also reply to posts with *comments*, and can attach comments to other comments as well. This allows social networks to be inferred, both at an inter-user level (Weninger, Zhu and Han 2013) as well as at an inter-community level (Olson and Neal 2013).

As will be demonstrated in this thesis, wear is of particular interest to researchers studying manipulation of content ranking in social media

sites. Voting on content submitted to a social media site is a form of edit wear: as users interact with the content submitted to the site, a number of them vote upon it, which is then used by the site software to determine the relative rankings of each piece of content. This ranking is an example of *algorithmic relevance*, which is discussed further in Section 2.3.7.

Hill et al.'s (1992) read wear and edit wear are used by Indratmo and Vassileva (2009) in their discussion of *social interaction history*, itself a form of edit wear. Social interaction history concerns itself with the analysis of interactions between humans within an information system: how often users comment, which users tend to be the most active, patterns of voting, and other related activity. This interaction history is edit wear: even if data within the system is not being modified, sufficient traces are left by the users of the system to generate usable edit wear. This concept of social interaction history was later developed into a more complete framework in Indratmo (2010).

2.3.6 Finding information in social media spaces

Social media spaces are a rich source of information, and benefit from the fact that the users of these spaces provide additional means for locating high-quality content. This section discusses the various ways in which past work has identified social means of finding information in spaces designed to support information retrieval with social elements.

Social navigation (Dieberger, Dourish, Höök, Resnick and Wexelblat 2000) is involved when people use information from other people to make decisions about navigation. Social navigation is not a novel concept, but rather as a technique that has been in use for as long as there have been humans interacting in any social information space. For example, recommendation engines, such as those used on shopping sites like Amazon.com

(Linden, Smith and York 2003) involve using information "left behind" by users in order to personalise an information space to suit the needs of the user.

Social navigation applies to both physical navigation, as well as to the navigation of information spaces. A frequently cited example of social navigation in the physical world is that of Svensson's (2000) "path in a forest": when many people walk through a forest over time, they provide "advice" to future walkers in the form of incrementally wearing a path into the ground. No explicit navigation aids are constructed, but rather the navigation advice is created as a secondary result of users navigating in the first place. This example has an immediate and direct link to the discussion of wear, presented in Section 2.3.5; in this case, the accumulated navigation advice takes the form of physical wear.

Socially-driven information navigation behaviours have taken significant metaphorical inspiration from physical navigation, and the underlying needs that drive physical navigation. Much of this extension of physical metaphor has been driven by the work of Pirolli; for example, the literature provides examples of information foraging and information diets (Pirolli and Card 1995, 1999), information scents (Pirolli 1997, Chi, Pirolli, Chen and Pitkow 2001); we also find information orienteering (O'Day and Jeffries 1993).

"Information foraging" is a theory described by Pirolli and Card (1995) that describes a tendency of information systems to evolve towards a stable state that maximises the gains of valuable information while minimising the cost of locating and making use of this information.

The analogy between information-seeking behaviour and food-seeking behaviour is echoed by the related concepts of *information diets* and *information scents*. Pirolli and Card's (1999) theory of information foraging de-

scribes information diets as the set of choices that an information seeker makes about which information sources to spend more time on than others. These decisions are informed by *information scents*: "residue" left by information that can be used to determine the perceived value and cost of the information, as it relates to the goal of the user.

These patterns of information consumption, in which users move from location source to location source in search of the higest value information, guided by their goals, relate closely to information orienteering, a term proposed by O'Day and Jeffries (1993) and extended by Teevan, Alvarado, Ackerman and Karger (2004), which describes a pattern of behaviour in which the user performs small, incremental searches as they narrow in on their goal. An information orienteering exercise involves broad-scale initial searches that provide additional constraints that allow the user to find the information they seek.

2.3.6.1 Exploratory searching

Exploratory searches are searches performed with nonspecific goals, which require analyses of multiple sets of information gathered over multiple iterations. When one searches for the date of Easter in a given year, that search is not exploratory, because a specific answer to a specific question is being sought; searching for information about the Apollo lunar landings (Wilford 1969) is exploratory, because no specific goal is in the searcher's mind when they begin looking into the general topic.

The browsing of a social media site fits the definition of an information exploration task (Bates 1989, O'Day and Jeffries 1993, Baldonado and Winograd 1997): a task in which the users look for new information within a defined conceptual area. Importantly, Baldonado and Winograd note that the conceptual area in which the user is searching for information

may be at any level of granularity; that is, a user may begin searching for new information about a specific topic, the field in which that topic is located, or even (at the higest level of granularity), any new information whatsoever.

When users engage in exploratory searches, they are uncertain about the specific information that they are looking for (White, Kules and Bederson 2005), but have enough of an understanding of the information they seek to be able to recognize when they have found something that fits their (vaguely-defined) criteria. White et al. (2005) notes that exploratory searching happens both intentionally as well as incidentally to other activities; in social media sites that cover a wide range of topics, the multidiscinplinary nature of the content provided creates a wide range of opportunities for serendipitous discovery of relevant topic areas, which sustain the searching behaviour.

This notion of exploratory searching builds upon previous work by O'Day and Jeffries (1993), which classifies searching behaviour into three *modes*: *a*) *following a plan*, in which users have a specific goal in mind and seek it out following a pre-planned search method; *b*) *monitoring*, in which users repeat the same search over a timespan, in order to note what results are new; and *c*) *exploration*, in which users follow an undirected exploratory path with no fixed goal in mind .

In the case of social media sites, the behaviour of users is a combination of both the exploration and monitoring modes: when users repeatedly visit a social media site, they are conducting an exploratory search with the goal of finding new results of that search. The search behaviour of social media site users is therefore a fusion of these modes, and may be considered equivalent to an unbounded, monitoring exploratory search.

Marchionini (2006) notes that exploratory searching is closely linked to

browsing behaviours; in particular, searches that are undertaken as part of a longer-term investigation mirror the incremental discovery of new information afforded by browsing.

2.3.7 Relevance

When users interact with any information retrieval system, they seek information that is *relevant* (Saracevic 1975, 2007, Cosijn and Ingwersen 2014). Saracevic (1975) notes that relevance is typically treated as an intuitive concept (Saracevic describes it as a "y'know"-like property, as if to suggest the following reply to the question of its definition: *"Well, it's what's relevant, y'know?"*).

Saracevic (1975) draws upon the seminal work of early information retrieval systems research, and presents an initial formal definition of relevance: a "measure of the effectiveness of a contact between a source and a destination in a communication process". This is a broad definition that may be applied to all forms of communication and the relationship between a source of information and a consumer of it, and serves as a useful starting point for pondering the question of what it means for information to be relevant to someone seeking information.

As a follow-up to his 1975 work, Saracevic (2007) presents a categorisation of different types of relevance, derived from the analysis of existing literature in the information systems field. Saracevic identified the following types of relevance, though added the caveat that this was not an exhaustive nor complete list:

- *System* or *algorithmic* relevance
- *Topical* or *subject* relevance
- *Cognitive* relevance or *pertinence*

- *Situational* relevance or *utility*
- *Affective* relevance

It is important to reinforce the fact that the breadth of scope in the discussion of relevance is deliberately constrained in this thesis to that which applies to the information-seeking behaviour of social media users alone. As with information spaces in general, users of a social media site rarely begin browsing the site with a particular goal in mind, but rather seek new content that matches their general interests at the time, and use what they find to drive further exploration of the information space (Baldonado and Winograd 1997).

Both Eslami, Aleyasen, Karahalios, Hamilton and Sandvig (2015) and Tufekci (2015) have done especially interesting work in examining the difference between content that an algorithmic relevance system (which is discussed in detail in Chapter 6) deems relevant, and that which a user would themselves deem relevant, while Gillespie (2014) presents a detailed discussion of the consequences of algorithms controlling and mediating the presentation of content on websites.

2.4 Social media ranking systems

Different types of ranking systems exist, which vary in their mechanisms. In this section, a selection of social media sites are reviewed, and their content ranking systems discussed. The structure and ranking systems of Slashdot, web forums, Digg and Reddit are each discussed.

2.4.1 Slashdot

Each comment on Slashdot has an associated score, which ranges from -1 to 5. When a comment is posted to Slashdot, its initial score is 1, if posted by a user, or 0, if posted by an anonymous user.

Periodically, users who are logged in are awarded a limited number of "moderator points", which are votes that they may assign to comments. A moderator may choose to expend these points however they wish, though they may only apply a single point to any comment. Moderator points (often abbreviated to "mod points") may be used to increase or decrease a comment's score by a single point. Additionally, when a moderation point is assigned to a comment, the user applying the point indicates the reason for the moderation, from a selection of adjectives, such as "insightful", "informative", "funny", and "overrated" (as seen in Figure 2.8). These selected adjectives are displayed to visitors to the site, with the intent that users are then able to modify their personal preferences on the site to give higher weightings to comments marked by moderators as "insightful".

Slashdot allows users of the site to moderate comments, but in a limited capacity. When a user's moderation points are expended, they cannot moderate. Moderation points cannot be transferred between users, and expire three days after they are awarded; users are urged to "use them or lose them", in order to prevent users stockpiling moderation points and using them only when they wish to heavily influence the discussion of a topic they feel strongly about (Malda 1999):

As Malda notes:

> *"I don't want people to stockpile their points. I want people to use them or lose them. Otherwise people will hold on to their X points until a story comes on that they have a strong opinion in, and they*

Figure 2.8: Moderation on Slashdot. Comments attached to posts have scores, as well as the selected reason for their high scores (as chosen by the moderator who voted the content up.)

will be tempted to moderate the discussion so as to sway things 'their way'."

In the Slashdot ranking system, individual moderation actions have heavy impact, due to their relative scarcity, and the fact that moderation points are a consumable resource (in that once a user has spent a point, they must wait until the system provides more to them, which may be some time).

2.4.2 Web forums

On a web forum, content is organised into threads, which are comprised of multiple posts. Different forum sites vary, but most order their threads such that threads with recent posts are displayed most prominently. The fact that the prominence of a piece of content depends on how recently that piece of content had a reply leads to some interesting behaviour from forum users; two of the more interesting behaviours are *thread bumping* and *thread resurrection*.

2.4.2.1 Thread bumping

Thread bumping is the practice of posting to a thread with primary intent of elevating its prominence on the forum, instead of contributing new discussion to the thread (Know Your Meme 2014). In web forums, the prominence of a thread is determined based on the date of the most recent post in that thread; this means that any post in a thread, no matter its content, will bring that thread up to the top of the page.

Consider the case of a user on a tech support forum who posts a question that is left unanswered. This user will see their post slowly move down the page as more active or more recently posted threads appear.

Most web forum software limits the number of threads displayed per page, which means that once a thread moves off the first page, the likelihood of other users seeing it (and therefore the likelihood of them replying) is considerably reduced.

Because the user wishes to have their question answered, they therefore want the thread that contains their question to be prominently displayed on the forum. They may post a new thread with the same question, but this may draw accusations of repeatedly asking the same questions that nobody has previously wished to answer. Another option available to the user is to post a reply to their *own* thread. The content of their reply does not matter; due to the rules of the forum software, the thread that has seen the most recent reply is the thread that is displayed most prominently.

Thread bumping, therefore, is so named because the thread is "bumped" to the top of the thread list without any additional content contributed to the discussion. It is generally seen as an annoyance to those other user who wish to see new content and discussion, rather than old content (Know Your Meme 2014). Many forums, therefore, have community rules that prohibit thread bumping.

It should also be noted that our example of an unanswered tech support question is not the only case in which thread bumping may occur; indeed, we may generalise our description of thread bumping behaviour to any case in which a user, dissatisfied with the prominence of a particular thread, causes that thread to become more prominent to the community at large without materially increasing the quality or appeal of the thread to the community.

2.4.2.2 Thread resurrection

A behaviour related to thread bumping is that of *thread resurrection*. This may be considered to be a variant on thread bumping.

In thread resurrection, a thread that has had a large amount of prior activity but has since lost prominence (due to the conversation tailing off) is brought back to the top of the list by a user posting a reply. The reasons for a user doing this vary: a user browsing old threads and noticing a comment that they wish to reply to may post a comment, resurrecting the thread; a user may also wish a thread resurrected simply because they enjoyed the earlier conversation. Threads are generally not resurrected by the user who originally posted them, but rather by other users who wish to continue the conversation.

Thread resurrection is generally considered to be poor etiquette, since there is a limited amount of space available on the first page of forum threads, and resurrecting an older thread pushes a more recent post off the first page and into the relative obscurity of later pages.

To that end, forums often have rules that prohibit thread resurrection. In some forums, such as Something Awful LLC (2014b), threads are automatically locked after a certain period of time after the last reply, preventing resurrection.

The Something Awful forum is a particularly interesting case, as the administrators of the site deliberately archive old but popular threads. To that end, Something Awful allows users to rate threads on a scale of 1 to 5, where 5 is high quality and 1 is low quality; threads that have a high average rating at the time that they are automatically locked are transferred to an area of the site called the "Comedy Goldmine" (Something Awful LLC 2014a), where they are available to read.

2.4.3 Digg

Digg (News.me Inc. 2012) was a "social news" site. Social news sites are social media sites that focus on both local and global news; instead of posting text, users posted links to other websites. The history of Digg, and its transformation from a user-driven social media site into an editorially-controlled news aggregator, is discussed in Section 2.2.4.

On Digg, users were able to post links, and reply to posted links in the form of comments. Unlike online forums, the prominence of a link was not entirely controlled by the recency of a post or of replies to that post, but is instead controlled by, among other factors, the number of "diggs" (votes) that a piece of content received. The specific algorithm used by the Digg website was not public. As a result, a cottage industry of users attempting to infer the algorithm and manipulate the rankings appeared Mezei (2006).

A user who is signed in was able to 'digg' a post, which applies a single vote to the story. Alternatively, a user may choose to "bury" the post, which applies a single *negative* vote to the story. The total of positive diggs and negative buries was then summed to create a score, which is then used in the calculation of the story's prominence.

2.4.4 Reddit

Reddit (Reddit Inc. 2012) is a social news site that uses a voting mechanism that influences voting. Users may post either links or text, and may comment on posts of either kind. Users may also vote on either links or comments. Votes may be positive or negative; the terminology used on Reddit for these are "upvotes" and "downvotes", respectively. An individual user may vote for each comment or story only once.

Reddit allows for several different modes of ranking content on the site. By default, content is ranked using an algorithm named "hot", which is discussed in detail by Salihefendic (2010). The hot algorithm combines the summed total of upvotes and downvotes, then weights the result based on the age of the post time: all other things being equal, older content is ranked lower than newer content. The prominence of a post on Reddit's site is therefore a measure of how quickly it amasses a large number of upvotes (while avoiding a large number of downvotes); once the post is popular enough to appear on the front page, it is exposed to many more users, who continue to vote it up.

This has the effect of keeping a popular link highly ranked, and therefore on the front page, for as long as there are people applying upvotes. As time progresses, the number of people who are both willing and able to give an upvote to a piece of content (remembering that a user may only vote for a post once) dwindles, and the effect of time outweighs the amount of votes a post has. The post's ranking quickly drops, and higher-rated, fresher content takes its place.

Reddit's "hot" algorithm has advantages over the simpler prominence method used on web forums. In an online forum, as is discussed in Section 2.2.2, the prominence of a piece of content is determined entirely based on the recency of activity; in the Reddit "hot" algorithm, a piece of content is prominent only for as long as enough users continue to vote it up (relative to all other content in the community). The problems of thread bumping and thread resurrection do not apply, since a single user cannot raise old content back to prominence on their own.

While "hot" is the default algorithm used to rank content on Reddit, it is not the only algorithm available. Others include "controversial", "new" and "top". These algorithms are considerably simpler than "hot": the

"new" algorithm ranks content based on how recently it was posted, and does not take into account voting or other activity; the "top" algorithm simply sorts based on votes, and takes nothing else into account; and the "controversial" algorithm holds content that has a large number of both downvotes and upvotes in high esteem. While it is very possible to choose a different ranking algorithm, most users do not use a different algorithm other than "hot".

The "hot" algorithm leads to its own effects on the behaviour of users. Because content that is posted to Reddit is guaranteed to fall off the front page at some point, a user who wants others to continue the discussion centered around a piece of content past the point when the community's combined interest is no longer sufficient to keep it on the front page cannot simply keep posting in the discussion thread. Rather, they must create a new post, and hope that it receives sufficient new interest to appear on the front page again.

2.4.4.1 Karma

Content on Reddit may be voted upon. When a piece of content receives a vote, its prominence is affected, in accordance with the "hot" ranking algorithm. In addition to the effect that it has on content, a vote also affects the user who posted that content.

On Reddit, each user has two score values, both termed "karma". Karma on reddit is divided into two categories: "link karma", and "comment karma". Link karma is the sum total of upvotes and downvotes received on content submitted to the site, while comment karma is the sum total of upvotes and downvotes received on comments posted by the user.

Karma does not affect ranking, and has no effect on the site beyond being displayed on a user's profile page. Reddit notes that karma is designed

to indicate "how much good the user has done for the Reddit community" (Reddit Inc. 2013).

However, simply by displaying the amount of karma that a user has, Reddit influences the posting behaviour of its users. Instead of the inquantifable benefit of "contributing to a good discussion", many users instead turn to posting in order to improve their karma score. Because a user's karma is dependent on the number of upvotes their comments and links receive, a user who wants to increase their karma should therefore post comments that are likely to be up-voted by other users.

In order to receive upvotes, a comment or link must be seen; in the case of comments, Weninger et al. (2013) note that Reddit users comment on the highest-scoring thread of comments, rather than the most topical. This increases their comment's visiblility, and increases the possibility that their content will be further up-voted (and thereby increase their karma.) Anderson, Huttenlocher, Kleinberg and Leskovec (2012a) note also that the temporal properties of a submission to a social media site (in their case, a question-and-answer site), strongly affect the eventual rating of a post, and thereby affect its visibilty to the community.

This practice can lead to users contributing content that is not necessarily novel or interesting, but is instead simply content that other users agree with. The sub-reddit dedicated to discussion of atheism, for example, has attracted criticiscm for being an echo-chamber - that is, a community that consists mostly of people agreeing with each other. This, in turn, has lead to the existence of the "circlejerk" reddit, which parodies this behaviour. This tendency towards self-aware parody is exemplified in this exchange, in a thread in which users were attempting to create a single sentence that would enrage the community the most (see Figure 2.9):

[User 1] Theirs *[sic]* alot *[sic]*, of ways to make redditors loose

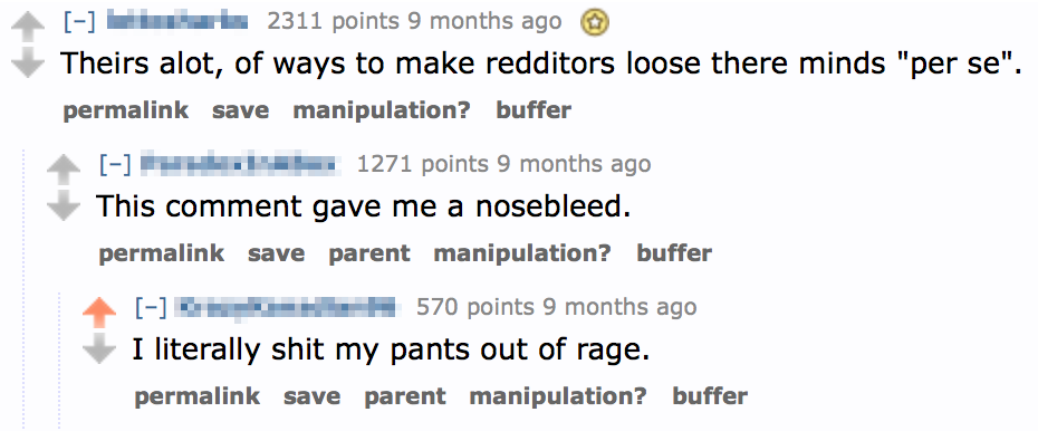


Figure 2.9: The exchange in question.

[sic] there *[sic]* minds "per se".[2]

[User 2] This comment gave me a nosebleed.[3]

[User 3] I literally shit my pants out of rage.[4]

2.5 Manipulation of social media sites

Any system that allows public participation may be subject to manipulation. One example of this is the phenomenon known as "astroturfing". Astroturfing is the creation of a fake "grassroots" campaign, in which multiple independent agents spread information in support of a cause (Cho, Martens, Kim and Rodrigue 2011). Ratkiewicz, Conover, Meiss, Gonçalves, Flammini and Menczer (2011) noted that it is possible to detect and trace the operation of an astroturfing campaign; while the area

[2] http://www.reddit.com/r/AskReddit/comments/1tbvm8/whats_one_sentence_that_will_absolutely_enrage/ce6fele, active as at October 24 2014

[3] http://www.reddit.com/r/AskReddit/comments/1tbvm8/whats_one_sentence_that_will_absolutely_enrage/ce6hwa0, active as at October 24 2014

[4] http://www.reddit.com/r/AskReddit/comments/1tbvm8/whats_one_sentence_that_will_absolutely_enrage/ce6mhky, active as at October 24 2014

of their study was limited to Twitter, this phenomenon can be generalised across other social media sites.

The manipulation of social media sites can be achieved by multiple individual users acting in a coordinated fashion. For example, as we have seen, content appears on the front page of Reddit's front page when it receives a large number of upvotes in a short period of time. This can be achieved naturally, by simply having the content be interesting and relevant to the user's interests (and thereby receiving the necessary number of votes), or it can be artificially induced by encouraging other users to upvote the content, regardless of their personal interest in the content.

Social media sites are not the only services that may be abused in this manner. The field of "search engine optimisation", or SEO, is centred upon causing search engine ranking systems to make specific content appear prominent to users than it would ordinarily have been (John et al. 2011). SEO is generally divided into "white hat" and "black hat" techniques: white hat techniques are centred on improving the content itself, in order to improve the legibility of the content and its accessiblity to the search engine system, while black hat techniques are designed to take advantages of the specific ranking algorithms employed by search engines, without increasing the quality or accessibility of the content itself.

Black hat SEO techniques are frequently changing, since search engines change their algorithms frequently. The reason for this is that a user accessing a search engine is after the highest-quality, most relevant content; content that appears prominently ranked due to black hat SEO techniques (and not because of its high quality) means that the user is likely to perceive the search engine as one that is not great at delivering good search results.

This is the reason why the specifics of Google's ranking algorithm are

a proprietary secret. If the details of the ranking algorithm were known, it would be significantly easier to engineer a black hat SEO technique that propels lower-quality content to a highly prominent position. As noted in Section 2.4.4, Reddit's ranking algorithm is publicly available; however, Reddit's spam detection system is not.

To quote one Reddit user:

> *"Tao of Reddit: The code that can be seen is not the true code.* [5]*"*

This means that while it is possible to engineer an exploit against the ranking system algorithm, such an attack must still deal with an unknown spam filtering system. (It is also worth considering the differing goals of a search engine and a social media site: when one searches Google for a topic, they often seek only a single, "best" result. On a social media site, users are not looking for a single they generally access several pieces of content; they are therefore looking for not a single "best" result, but several "good" results.)

The effort involved in manipulating social media networks may be delegated to outside agents: Motoyama, McCoy, Levchenko, Savage and Voelker (2011) use the term "abuse work" to describe outsourced freelance abuse and manipulation of web services. In their example case of search engine ranking manipulation, this includes account creation, social networking link generation and search engine optimization support. In the case of Reddit, applicable forms of abuse work include the creation of multiple user accounts, and upvoting or downvoting content. This behaviour is explicitly banned by Reddit's rules for users, which prohibit "voting rings": groups of users who focus their voting in a coordinated fashion. (Reddit Inc. 2013)

[5] http://www.reddit.com/r/WTF/comments/eaqnf/pardon_me_but_5000_downvotes_wtf_is_worldnews_for/c16qemu, active as at October 24 2014

Posts that take into account the social aspects of social media sites can have real-world impacts. Althoff, Danescu-Niculescu-Mizil and Jurafsky (2014) discusses the Reddit community "Random Acts of Pizza" (Morrison 2014), in which users donate pizza deliveries to others in the community who are in need.

2.5.1 User suspicions of social media site manipulation

To illustrate manipulation in more concrete terms, it is useful to present and discuss an incident involving manipulation that took place on Reddit in May 2012. Links to each of the comments quoted in this section appear as footnotes.

Reddit, as is discussed in further detail in Section 2.4.4, is comprised of multiple sub-communities, known as *subreddits*. Different communities have different focuses: the communities */r/news* focuses on world news, while */r/videos* focuses on interesting videos, generally hosted on external sites such as YouTube (YouTube LLC 2014)

On the 4th of May 2012, a Reddit user posted a video to the */r/videos* subreddit, featuring a soldier arriving home from deployment in Kuwait and being greeted by his very emotional family.

In the comments section of this post, a user commented that they had noticed what they believed to be a trend of unusual posting behaviour involving similar videos of soldiers returning home[6]:

> *"I feel like I'm the only one who notices this, but... I find this strange... not the video - the video is touching. But every few times a month a "Welcome Home Blog" video gets posted, hits the front page*

[6] http://www.reddit.com/r/videos/comments/t6pqc/man_absolutely_floored_by_the_return_of_his/, active as at October 24 2014

and it's always by an account that this is the singular submission. Then the person deletes the post and their account. [...] I mean... are we a part of some sort of experiment? It's just strange, man.[7]"

This prompted a discussion regarding the posting histories of the user who had originally posted the video, the user who had the top-ranking comment on the poster, and the user who had to top-ranking reply to that comment. Usernames in the following passages have been replaced with numbered indicators.

"*Dig a little deeper and it gets weirder... the submitter: redditor* [User 1] - *1 month, this is his only submission. top comment from* [User 2] - *redditor for only 1 month, her comment is her only comment ever. the person who responded to* [User 2] *(*[User 3]*) and has the next largest amount of votes... redditor for a month, her comment is her only comment ever.* [8]"

Other users then began chiming in, noting that these three users all appeared to have joined reddit at the same time.

"*Even more... although* [User 1] *joined 4/5/12,* [User 2] *and* [User 3] *joined on the same day.* [9]"

"*Not only the same day, they joined within three minutes of one another. March 27 15:07:28 and March 27 15:10:33.* [10]"

[7] http://www.reddit.com/r/videos/comments/t6pqc/man_absolutely_floored_by_the_return_of_his/c4k329k, active as at October 24 2014

[8] http://www.reddit.com/r/videos/comments/t6pqc/man_absolutely_floored_by_the_return_of_his/c4k5mry, active as at October 24 2014

[9] http://www.reddit.com/r/videos/comments/t6pqc/man_absolutely_floored_by_the_return_of_his/c4k9bov, active as at October 24 2014

[10] http://www.reddit.com/r/videos/comments/t6pqc/man_absolutely_floored_by_the_return_of_his/c4kbhtn, active as at October 24 2014

> *"Here is the* [link to the user page] *for* [User 2]. *Their only post is one congratulating new enlistments, 2 years ago.* [11] *"*

This discussion serves as an example of users being suspicious of, but not being able to prove, content manipulation on the site: in this case, possible content manipulation potentially aiming to associate positive emotions with the military.

It is not the goal of this thesis to make specific accusations of manipulation. However, identifying specific cases of users indicating their suspicions allowed the author to develop an understanding of what users view as manipulation, and their sentiments towards it. Indeed, these observations of suspected manipulation incidents was instrumental the basis of the first phase of research (presented in Chapter 3), in which administrators on Reddit were interviewed on whether they had seen anything they viewed as potential examples of manipulation.

2.6 Conclusions

The review of past research presented in this chapter has established the context necessary for the discussion of the research presented in this thesis. Several factors critical to the understanding of this research and the contributions that it makes to the literature have been presented and discussed, which allowed for the development of specific objectives for the research. In addition, the agenda and scope of this research has been sufficiently constrained so as to allow for useful contributions to be made by the research.

[11] http://www.reddit.com/r/videos/comments/t6pqc/man_absolutely_floored_by_the_return_of_his/c4kcp3d, active as at October 24 2014

3

Phase 1: Is It Really A Problem?

This chapter reports on interviews with administrators and moderators at Reddit and an analysis of community guidelines used by a number of communities. In doing so, an understanding of the various different types of manipulation encountered by moderators was built, which was used in the phases following the one discussed in this chapter.

3.1 Introduction

The goal of this research was to study manipulation in social media sites. This required gaining an understanding of:

- whether manipulation took place, and if it did, what administrators and users thought manipulation *was*; and
- what the impact of this manipulation was.

The first and second phases, presented in Chapters 3 and 4, answer the first point. The third phase, presented in Chapter 5, answers the second.

The purpose of the first phase of the project was to establish an empirical grounds for the research, driven by the need to create a better understanding of what manipulation actually *is*, in the context of the research. An a priori definition of manipulation is exceedingly difficult to establish, especially when one lacks an understanding of the context in which that manipulation is taking place. This first phase, therefore, sought to establish this context, in order to allow for a more detailed exploration in successive phases.

Specifically, this first phase sought to find answers to two critical questions: first, whether manipulation was present on Reddit, and second, what forms of manipulation were being observed.

In order to do this, two sources of data were consulted. First, interviews with moderators and administrators at Reddit were conducted, in order to gather perspectives from individuals whose role at the site is to ensure a high level of quality in the content that appears on the site. Second, the text of the community guidelines and rules used by a selection of popular subreddits was collected, and analysed alongside the transcripts of the administrator and moderator interviews.

3.1.1 Chapter structure

The presentation and discussion of the first phase of the research is conducted over the following sections:

- Section 3.2 provides an overview of the study, and the objectives served by it.
- Section 3.3 reports in further detail on the first aspect of the study, which comprises interviews with site administrators. The design

and conduct of the interviews is presented, including the methodology, ethics, and analysis of the resulting data.

- Section 3.4 reports on the second aspect of the study: namely, the analysis of subreddit rules. The design of the study, including the choice of methodology, and selection process, are presented.
- Section 3.5 reports on the analysis of the collected data from both components.
- Section 3.6 interprets the findings from the analysis, and presents an initial description of the different types of manipulation identified in this study.
- Section 3.7 discusses the implications of the findings for the following studies conducted in this research.
- Section 3.8 summarises the work reported on in this chapter.

3.2 Approach

The first phase of the research involved determining what site administrators considered to be the manipulation of the ranking system used on the site. This served to confirm the research focus, and determine how the two main classes of users on the site perceived content manipulation. This data was reinforced by the collection and analysis of community guidelines on subreddits.

The structure used in this section is as follows:

- Section 3.2.1 discusses the objectives of this study.
- Section 3.2.2 discusses the specific ethical considerations for the interview component of the study.

- Section 3.2.3 summarises the contributions made by this phase to the overall research presented in this thesis.
- Section 3.2.4 discusses the philosophy behind the research, and the consequences of this philosophy on the metholodgy.

3.2.1 Objectives

The first phase of the research was designed to provide an empirical basis for understanding manipulation, in order to allow for successive phases - described in Chapter 4 and Chapter 5 - to begin to be addressed. In doing so, RQ1 (from Section 1.2), which asked what the most prevalent types of manipulation that exist on Reddit, could begin to be answered.

The specific objectives of this study were to:

1. *establish whether administrators and moderators consider manipulation to be present on Reddit*; and to
2. *identify the different types of manipulation observed by administrators and moderators.*

These objectives serve the wider objectives of the research, presented in Section 1.2, by establishing a context for the discussion of manipulation.

3.2.2 Ethics

The study reported on in this chapter was approved as a Minimal Risk Study by the Tasmanian Human Research Ethics Committee. The reference number for this phase of the study was H13025. Precautions around data collection, handling and analysis of the interview transcripts were taken, due to the potentially private and sensitive nature of the conversations. These precautions included:

- Participants were not required, at any time, to disclose any personally identifying information about themselves or any other moderators on Reddit.
- Participants were free, at any time, to withdraw from the study at any time prior to the completion of data collection.
- Analysis of the interviews was carried out after identifying information was stripped from the data.
- Participants were comprehensively informed that no judgements were being made on their individual roles as moderators.

3.2.3 Contributions

This chapter makes the following contributions to the overall thesis:

1. *The establishment that manipulation exists, and is viewed as a problem by administrators and moderators.* This was critical to the fundamental goal of the research, and enabled the second and third phases to continue as planned.
2. *The identification and classification of the different types of manipulation observed by moderators and administrators on Reddit.* The data from this study is used in the second and third phases of the research as the foundation for categorising various types of manipulation in the second phase of the research, which is repoted on in Chapter 4.
3. *The identification of several novel, previously-unknown forms of manipulation.* Several forms of manipulation have been previously identified in past research, such as the Sybil attack (Douceur 2002). However,

the findings presented in this chapter indicate the existence of several previously-unidentified methods of manipulating the ranking of content in social media sites, including the "shaming bot" discussed in Section 3.6.8.

3.2.4 Research Philosophy

When presenting a body of research, it is necessary to first discuss the underlying philosophy of knowledge that underpins the researcher's approach to gathering that knowledge. This is because the variety of different epistemologies, ontologies and models of human nature incline researchers to different choices of methodologies among social scientists (Burrell and Morgan 1979).

Researchers therefore need to identify and explicitly declare their alignment to which of the many ontological and epistemological positions, and of their models of human nature, in order to provide the background against which their research is conducted.

Without this background, a reader who subscribes to a different epistemological position lacks the context against which the conduct and interpretation of the research takes place, and could reasonably assume that, lacking other information, their own positions apply; this can create a situation in which the reader interprets the thesis in a different manner to that which the writer intended.

In order to situate this thesis, this section presents the positions of the author.

Guba and Lincoln (1994) state that the underlying beliefs that define paradigms of inquiry can be summarised by answering three questions:

- What is the form and nature of reality and, therefore, what is there

that can be known about it? (Ontology)

- What is the nature of the relationship between the knower or would-be knower and what can be known? (Epistemology)
- How can the inquirer go about finding out whatever he or she believes can be known? (Methodology)

These questions are connected: the answer to any of these questions constrains the possible answers of the rest.

3.2.4.1 Ontology

Orlikowski and Baroudi (1991) discuss ontology as the empirical world being either *objective* or *subjective*. Objective ontology describes the empirical world as independent of humans, while subjective ontology describes the empirical world as a result of the actions of humans.

This research is an exploratory study into the results of humans interacting within a social space, and, as a result, the researcher acknowledges a subjective ontology which "focuses on the meanings that people give to their environment." (May 2011).

The research aims to investigate the manipulation of social media sites, and the results of that manipulation upon both the social media site and upon the people who participate in that site. Both of these items must necessarily be considered from a subjective position, because users in a social media site react to the content of a social media site in their own subjective ways. Therefore, this research is directly concerned with the different meanings that people in these sites attach to the content that they are reading.

3.2.4.2 Epistemology

Two important, yet contrasting positions of epistemology are positivism and interpretivism. Positivists hold that knowledge is only valid when derived from observational data, including sociological knowledge. Interpretivists hold that social topics cannot be understood unless this subjective creation of meaning is taken into account; interpretivist social scientists see social reality as created out of human participation and interpretation. Orlikowski and Baroudi (1991) note that interpretivists believe that people create their own, subjective meaning as a result of their interactions with the world around them.

Social reality, therefore, is the result of humans subjectively interpreting their interactions with other humans, and their own evaluations of these interactions, in a social environment. In the context of social media, the entire content of social media sites is the result of people interacting with each other through a technologically mediated system.

When one considers the possibility of manipulation occuring within this system, it is an inescapable conclusion that both the commission and observation of this manipulation can only be discussed in the context of the subjective experiences of people who encounter it. It is not possible to consider manipulation as an objectively measurable force that is identifiable without a human to observe and identify it; it is the direct result of humans interacting within a social media system.

As a result, this thesis contends that the only valid epistemological position for this topic is an interpretivist one. This has consequences for the selection of the methodology of this thesis; open coding, in the context of the grounded theory approaches discussed in Section 3.5, supports the identification of themes with an interpretivist approach.

3.3 Data collection: Administrator interviews

Reddit was used as the case study for the research into manipulation. The specific selection of Reddit as case study was based on the fact that Reddit has a large user population, combined with a very broad array of different topics. This makes Reddit into a microcosm of different interests, and makes it extremely interesting to study.

Case studies are deep dives into a specific instance, person, or event, with a view to creating an understanding of phenomena and its related causes (Yin 2003). A case study was the most effective means of researching manipulation, due to the fact that most social media sites have a subtly different structure; it was felt that separating the differences in structure across a large collection of sites would have interfered with data collection.

The first source of data used in this study came from semi-structured interviews conducted with administrators and moderators of Reddit. This section reports on the design, conduct and analysis of the interviews in further detail. The objectives of these interviews were discussed in Section 3.2.1; the findings from the interview are discussed in Section 3.6.

The structure of this section is as follows:

- Section 3.3.1 discusses the choice of methodology used.
- Section 3.3.2 discusses the design of the interviews.
- Section 3.3.3 discusses how participants were recruited.

The discussion of the analysis of the interviews is discussed concurrently with that of the analysis of the community guidelines, in Section 3.4.

3.3.1 Interviews

In this first phase of research, information from both administrators and moderators was sought regarding their thoughts on content manipulation on Reddit in order to address the objectives presented in Section 3.2.1.

As discussed in Section 3.2.4, interviews are an appropriate method of gathering data when the one's approach to the research is an interpretivist one. Interviews require their participants to reflect on their experience and provide their interpretation as part of the context of the data they provide.

Interviews were also selected as the method of choice for this research due to their ability to deeply examine real-world behaviour in natural settings (Drever 1995), and to collect detailed information that would be otherwise challenging to gather. Interviews allow users to reflect and consider what they are talking about, which is a feature not captured by other means of data collection such as questionnaires (Lazar, Feng and Hochheiser 2010).

Interviews have their drawbacks; the amount of time needed to meaningfully conduct an interview with a single subject and to transcribe that data is significant. However, if unbounded conversations can be managed, interviews provide a great deal of flexibility for the researcher (Robson 2002, Lazar et al. 2010).

Interviews are frequently combined with other techniques for collecting data, as this helps the researcher determine the relationship between behaviours and perceptions (Crabtree and Miller 1999). This was done during the research discussed in this thesis; the interviews conducted during Phase 1 of the research, discussed in this chapter, was reinforced by the web-based data collection conducted in Phase 2 (discussed in Chapter 4), the data from which was in turn further explored by more interviews in

Phase 3 (discussed in Chapter 5).

The interviews with administrators were conducted in a *semi-structured* fashion. In a semi-structured interview, the interviewer does not have a fixed set of questions, to which he or she writes down the answer for each question asked; rather, the interview is defined as a set of pre-defined *focus areas*, in which the interviewer is free to ask relevant questions (Drever 1995).

A thematic approach was used in the analysis of both the interviews and the community guidelines. The thematic approach for this analysis takes significant inspiration from the techniques used in grounded theory (Braun and Clarke 2006), which is suited for the analysis of early components of work without the researcher having to commit themselves to using the entire suite of methods and frameworks involved in grounded theory, as described by Corbin and Strauss (1990).

In this study, the transcripts of the conducted interviews were analysed concurrently with the text of the community guidelines. Because both sources of data applied to the same topic, both sets of data were able to be analysed side-by-side. This was found to assist in the analysis of both, in that themes that existed in one were found to exist in the other, which served to validate their inclusion in the final analysis.

3.3.1.1 Semi-structured interviews

Semi-structured interviews are powerful and flexible tools for data collection in cases where specific questions that need asking may not be known until part-way through the interview. Drever (1995) notes that this form of interviewing combines the flexibility of discussion, with the option for rigidity when it is needed: an interviewer is free to explore a focus area with the subject as far as is useful, and is able to move to other areas when

necessary. While conducting these interviews, the author found that this observation to accurate: conversations remained usefully on-topic, while still allowing for flexibility and exploration of related issues.

Human-computer interaction is a field in which semi-structured interviews have been used regularly (Robson 2002); as Kjeldskov and Graham (2003) note, questionnaires and interviews are "respected and widely used instruments" for the collection of data in this area. Interviews, like other survey-based research techniques, are useful in gathering information about user experience; given that the topic of this research is about the experience of users with regards to content manipulation, interviews are a particularly appropriate data-gathering tool.

Consequently, semi-structured interviews were selected for the first component of the data collection confudcted in this phase of the research as they simultaneously allow for great flexibility in data gathering, as well as rigidity when needed (Drever 1995). The use of semi-structured interviews is well-established in qualitative research (Robson 2002), and has proven useful in the study of online communities (Konstan and Chen 2007).

Semi-structured interviews were appropriate for this phase of the research in particular because of the fact that they permit the interviewer to explore aspects of the area under discussion in directions that are not known at the time that the interview questions are devised. This was particularly useful in the case of this study, as they allowed for an exploration of a topic whose definition was in the process of being understood.

As Louise Barriball and While (1994) notes, semi-structured interviews allow for a close examination of people and their working situations; as social media site moderation can be considered a form of (largely volunteer) 'work'; additionally, semi-structured interviews are an extremely so-

cial form of gathering data (Robson 2002), which makes them appropriate for gathering information about an especially social field of study.

3.3.2 Design

The interviews conducted during this phase of the study focused on the role of moderators on the site, their perspectives on manipulation, and on the different types of manipulation noticed by the participants. The role of an administrator or moderator on a social media site is to ensure that high-quality content is promoted, while low-quality content is demoted or removed; in order to achieve these goals of content quality, these users (also referred to as owners or hosts) have the power of allowing or rejecting posting, removing users from a community, and have a higher level of control over the content of the site (Butler, Sproull, Kiesler and Kraut 2007). However, to ensure that the scope of this research remained effectively constrained, all conversation was deliberately limited to the discussion of manipulation.

Specifically, when the word "manipulation" first began being discussed in the interviews, the interviewer clarified that "manipulation" referred to attempts to influence the ranking of content. Other forms of anti-social behaviour, including bullying, personal attacks (including the use of racist, sexist or other exclusionary language), or attempts to reduce the availability of the site (for example, denial-of-service attacks, as per Mirkovic, Dietrich, Dittrich and Reiher 2004) were not considered, as these are not attempts to manipulate the ranking of content, but attempts to stir up discussion or express an unpopular opinion (rather than attempts to increase or decrease the ranking of submitted content).

Background questions used in the interview were used to establish an overview of the role that each participant played in the moderation of

communities on reddit, including their history, how long they have been moderating, and so on. Subsequently, focus questions were used to elicit specific kinds of data from participants, and were used to drive the conversation towards areas of interest to the research. The general topics for the interviews were derived from the literature discussed in Chapter 2.

The background questions used in the interviews included:

- How long have you been a moderator?
- Which subreddits do you moderate?
- How active are you in moderating these subreddits?

Examples of the focus questions used include:

- Do people attempt to manipulate the ranking system in the subreddits that you moderate?
- What different kinds of manipulation do you see?
- What processes do you have in place for determining whether a user is manipulating a subreddit?
- How did the code of conduct that you have on the subreddits that you moderate evolve?

Interviews with administrators were conducted face-to-face. This was made possible due to a fortunate coincidence of the author being present in San Francisco for a conference; a face-to-face meeting was quickly arranged, and took place over the course of an afternoon. This opportunity was extremely useful, as it allowed for rapid collection of more data than would have otherwise been possible over the original plan of interviews over Skype.

Interviews with moderators were conducted via Google Talk, a text-based instant messaging system. This allowed for automatic transcription of the interviews, which was useful for several reasons:

- Errors in transcription were eliminated, as all conversation was directly logged in its original format
- The full context and phrasing of participants was preserved, avoiding potential accidental omissions
- Anonymity of the participants was enhanced, due to the fact that no audio recordings needed to be made

Following the completion of the design of the study, recruitment of participants began.

3.3.3 Recruitment and participation

Participants were recruited from both the administrative staff of Reddit, and from moderators of subreddits. Administrators at Reddit were already known personally to the author, which allowed email invitations to be sent directly to them. Interviews with the staff took place in the Reddit offices in San Francisco during March 2013; two administrators participated in face-to-face interviews.

Additionally, invitations to participate in the first phase of the research were sent to all moderators of a selection of high-population subreddits. The specific subreddits that received invitations were:

- /r/pics
- /r/funny
- /r/AskReddit

- /r/worldnews
- /r/todayilearned
- /r/science
- /r/IAmA
- /r/WTF

This selection of subreddits was taken from the list of subreddits ranked by subscriber count, as calculated by the third-party Reddit statistics site "Stattit" (Birch 2013). The process for this selection was by taking the top ten subreddits listed; it was noted that two of these subreddits are used only by administrators to post site-wide announcements, and restrict non-administrators from posting any content; because these do not allow the general users to post user-generated content, they are unsuitable for the discussion of user-generated content.

Of the eight subreddit-wide invitations submitted, two subreddit moderators replied. The semi-structured interviews were therefore conducted with four participants: two subreddit moderators, and two Reddit administrators.

This number was sufficient for useful data collection in this survey, because each participant had a ground-level view of the situation inside the social media site. It must be acknowledged that this number of semi-structured interview participants is not necessarily a representative sample, and thus are not statistically significant. The results should be interpreted as suggestive, rather than providing conclusive findings: they form a basis for future components of this research. Additionally, participant observations were reinforced by the data gathered from the community sub-reddit rules, discussed in Section 3.4.

3.4 Data collection: Sub-reddit rules

This section discusses the second component of the data collection, in which the community guidelines used in a selection of subreddits were collected and analysed following the analysis of the interviews described in Section 3.3.

The subreddit rules are the source of instruction for both users who post to a subreddit and the moderators of that subreddit. To that end, subreddit rules can be seen as the accumulation of community norms, specific to each subreddit, as well a reflection of the positions of community moderators. Subreddit rules, therefore, serve as an important complementary source of data in the study of manipulation.

The structure of this section is as follows:

- Section 3.4.1 discusses the selection of the subreddits whose subreddit rules were analysed.
- Section 3.4.2 discusses the nature and content of the selected subreddits.

3.4.1 Selected Subreddits

The selected subreddits included the top 10 subreddits as listed on stattit.com (see Figure 3.1, ordered by number of subscribers, but did not include admin-only subreddits (which do not allow users to add new posts, but rather display posts only from Reddit administrators.)

This selection was:

- /r/funny
- /r/pics

- /r/AskReddit
- /r/todayilearned
- /r/science
- /r/IAmA
- /r/WTF

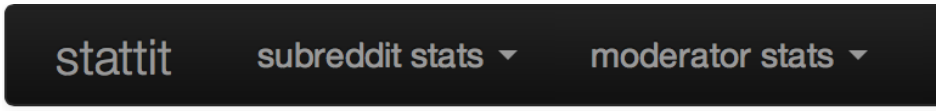
The methods used in the research presented in this thesis selected individual subreddits on the basis of the number of subscribers. This selection took place prior to the publication of Olson and Neal (2013); however, the results published in the subreddit map presented by Olson and Neal indicated that selected subreddits used in this research had a high degree of connection to other subreddits.

In addition to this selection from the top ten subreddits, a collection of other subreddits with a lower ranking was selected, which allowed the analysis to consider much more topic-specific subreddits than the top reddits, which tend to cover much broader subject areas.

Selected lower-ranking subreddits were chosen from the list of subreddits available at Reddit's list of subreddits[1], which met the following criteria:

- *The subreddit has an active community, and at the time of writing had submissions posted within the last week.* This requirement ensured that the rules of the subreddit were reasonably up-to-date.
- *The topic of the subreddit was one that the researcher had at least passing knowledge of.* This requirement ensured that the analysis of the subreddit's rules could be done in the context of the topic: any lack of

[1] http://www.reddit.com/subreddits/, active as at October 24 2014



Subscribers - full list

#	Subreddit	Subscribers
1	/r/funny	3,921,497
2	/r/pics	3,845,289
3	/r/AskReddit	3,673,683
4	/r/todayilearned	3,536,067
5	/r/worldnews	3,514,004
6	/r/science	3,446,305
7	/r/announcements	3,442,327
8	/r/IAmA	3,398,333
9	/r/blog	3,392,811
10	/r/WTF	3,334,493

Figure 3.1: The top 10 subreddits, ranked by subscriber count on analytics site stattit.com (Birch 2013)

understanding of important terms used in the community could be avoided, and misinterpretations were mitigated or eliminated.

- *The topic of at least several subreddits were ones that the researcher had no direct involvement with.* For example, while the author is able to comprehend a discussion of rules surrounding tattoos, the author is not a member of any tattoo-related social media community. This prevented the set of additional subreddits from being dominated with subreddits familiar to the author.

- *The subreddit is not marked as "NSFW" ["not safe for work"]*: Subreddits that contained adult material, such as pornography, were not considered for inclusion.[2]

The final list of additional subreddits, in addition to those selected from the top ten, were:

- /r/explainlikeimfive
- /r/YouShouldKnow
- /r/pokemon
- /r/GameOfThrones
- /r/Skyrim
- /r/tattoos
- /r/circlejerk
- /r/photoshopbattles

[2]The author has no specific objection to this content, but felt that it would distract from the subject at hand.

- /r/conspiracy

For the most part, subreddits list their submission rules as part of the main subreddit rules, which are shown in the side-bar of the site. Some subreddits diplay their submission rules in separate pages, linked to from the sidebar; in those cases, the text of the submission rules were used for analysis as well.

3.4.2 Subreddit content

This section discusses the nature and general content of the subreddits selected for analysis of their community guidelines.

3.4.2.1 /r/funny

/r/funny focuses on humourous content, of any format. Generally, this takes the form of still or animated pictures.

While the description of the subreddit reads *"You may only post if you are funny."*, the subreddit rules of /r/funny specifically forbid certain types of content, and direct users who wish to post these kinds of content to other, more specialised subreddits. Examples include *"no reaction gifs or HIFW* ["how I feel when"] *posts; no pictures of just text; no DAE* ["does anyone else"] *posts."*

3.4.2.2 /r/pics

/r/pics describes itself as "a place to share interesting pictures". Content submitted to /r/pics is required to be a still, non-animated picture.

Like /r/funny, the overall description of the subreddit is quite broad: *"A place to share photographs and pictures."*. However, the description goes

on to state: *"note that we are not a catch-all for general images (of screenshots, comics, etc.)"*

Accordingly, the subreddit rules include specific restrictions on certain types of post, including nudity, gore, personal information, or images with superimposed text. Several competing image-based subreddits exist, including /r/AnythingGoesPics, which encourages all submissions.

3.4.2.3 /r/AskReddit

/r/AskReddit is a platform that allows users to pose questions to the community at large, in the aim of creating an interesting discussion. Submissions are required to be text-based (that is, no links to other sites, and no links to media like pictures or videos.)

Questions are open-ended, though are sometimes aimed at specific demographics. Examples include *"Older people of Reddit, what is something that was something you never thought was possible, but is available today?*[3]*"* and *"What fads of the 2010s will be ridiculed in the 2020s?*[4]*"*

3.4.2.4 /r/todayilearned

/r/todayilearned is a subreddit that focuses on interesting facts. Users post a summary of the fact, and link to additional information. Posts to this subreddit are required to begin with "TIL", an abbreviation for "Today I Learned".

An example of this format is *"TIL in Ancient Persia, the men used to debate ideas once sober and once drunk, because the idea needed to sound good in both*

[3] http://www.reddit.com/r/AskReddit/comments/1joe8t/older_people_of_reddit_what_is_something_that_was/, active as at October 24 2014

[4] http://redd.it/1jojbs, active as at October 24 2014

states in order to be considered a good idea.[5]"

3.4.2.5 /r/science

/r/science is a community for discussing recent research. Posts are required to be either direct links to papers, or to scientific news sites discussing recent research, and editorialising is not permitted.

Posts submitted to /r/science are annotated with the field of research - for example, environmental science, medicine, astronomy, and so on. This allows visitors interested in particular fields to choose to only see posts relevant to their interests.

3.4.2.6 /r/IAmA

/r/IAmA is a "public interview" platform, in which users who believe they have an interesting history, profession or knowledge to discuss make themselves available for public questions.

/r/IAmA has become known as a platform for media personalities and celebrities, who often use /r/IAmA as a promotional tool, due to the public and open-ended nature of the discussions that take place there. Some of the more well-known discussions involved President Barack Obama, Bill Gates, and Arnold Schwarzenegger; however, non-celebrity users are equally popular and more commonplace. At the time of writing, one of the top posts on the page was *"I am a wind turbine technician. AMAA* [ask me almost anything].[6]"

[5] http://www.reddit.com/r/todayilearned/comments/1jnyow/til_in_ancient_persia_the_men_used_to_debate/; the user is referencing Herodotus (430 BCE)

[6] http://www.reddit.com/r/IAmA/comments/1jmm8j/iama_wind_turbine_technician_amaa/, active as at October 24 2014

3.4.2.7 /r/WTF

/r/WTF is an area for users to post disturbing, curious, surprising or otherwise interesting content that is not suitable for general consumption.

Content submitted to /r/WTF frequently contain gore and pornography, although the ostensible intent of the community is not to arouse. As a result, /r/WTF is one of only a few subreddits that does not show preview thumbnail images of content.

3.4.2.8 /r/explainlikeimfive

/r/explainlikeimfive is a community for providing simple answers to complex topics or questions. Users post requests for explanations of topics, such as *"Why aren't people buying the $1 houses in Detroit?"* or *"Explain like I'm 5: Different types of assets: shares, bonds, options, funds and derivatives."*

While answers are not expected to be written as though for people who are actually five years old (the subreddit rules note that *"preschooler-friendly stories tend to be more confusing and patronizing"*), the subreddit expects that users take care to explain things as simply and clearly as possible.

3.4.2.9 /r/YouShouldKnow

/r/YouShouldKnow is a location for users to share *"obscure things that most should already be aware of, but aren't."* Content posted by users generally takes the form of links to resources, facts, or other information that users consider worth knowing about.

The subreddit focuses on self-education and general advice; in terms of content, it shares a common area of focus with /r/todayilearned, though /r/todayilearned users generally prefer specific facts, rather than advice.

3.4.2.10 /r/pokemon

/r/pokemon is a subreddit for discussing the Pokémon media franchise created by Nintendo (Nintendo Inc. 2014). Posts are generally links to Pokémon-related images, with occasional links to articles discussing the Pokémon games.

/r/pokemon specifically bans, among other content, *"anything unrelated to Pokémon"* and *"links to or requests for ROMs* [pirated copies of the Pokémon games]".

3.4.2.11 /r/GameOfThrones

/r/GameOfThrones is a discussion community for *Game of Thrones*, the television show based on "*A Song Of Ice And Fire*", a series of fantasy books written by George R. R. Martin (1996).

Like many episodic stories, this series of books and televisions shows is prone to "spoilers" - that is, premature disclosure of plot information. To prevent community members from inadvertently exposing themselves to spoilers, /r/GameOfThrones implemented a system of tags and labels for content submitted to the subreddit: users who wish to discuss the story consequences of an event that takes place at a particular point in the story are able to label their submission as containing "spoilers" for that point. This allows people who have not yet reached that point in the story to browse the rest of the community without accidentally exposing themselves to spoilers.

3.4.2.12 /r/Skyrim

/r/Skyrim is a discussion community for the video game "Skyrim" (Bethesda Softworks LLC 2014). Posts in this subreddit generally involve discussion

of the plot, sharing of screenshots, and discussions of user-created modifications to the game.

However, the /r/skyrim subreddit prohibits (among other content) images with superimposed text. This prohibition is shared with /r/pics, /r/pokemon and /r/GameOfThrones.

3.4.2.13 /r/tattoos

/r/tattoos is a discussion community for tattoos. Posts to this subreddit are generally photographs of tattoos, and discussions of same; the subreddit specifically prohibits offers of sales of goods or services.

The community also prohibits any kind of discussion relating to the pricing of tattoos, in an effort to keep the discussion focused on the artistic merits of the tattoos themselves.

3.4.2.14 /r/circlejerk

/r/circlejerk is a subreddit that exists to satirise Reddit as a whole. Posts to this subreddit generally parody current trends on Reddit, and make fun of phrasing commonly used in the manipulation of ranking of content on Redit:

> *"I told my friend that /r/circlejerk Would Upvote a Sock to the Front Page, And He Said There Was No Way*[7]*."*

3.4.2.15 /r/photoshopbattles

/r/photoshopbattles is a "creative battle forum", in which users post an innocuous image, and other users post modified versions for humourous

[7] http://redd.it/1tmszs, active as at October 24 2014

effect. For example, a user may post a picture of a kitten popping a balloon[8]; this was then modified such that the kitten was in fact playing guitar at a rock concert[9].

3.4.2.16 /r/conspiracy

/r/conspiracy is a forum for discussing conspiracy theories.

Posts generally relate to news and developments that community members find suspicious; the subreddit rules for this community are deliberately lenient, and focus mostly on poster behaviour rather than post content (for example, *"Derisive slurs against people's race, religion, ethnicity, nationality, social order or creed are not tolerated."*).

3.5 Analysis

On its own, data is of very limited utility to researchers. Once gathered, it must be analysed in order to produce useful results. Much of the data analysis performed in the social sciences is performed using techniques derived from *Grounded Theory* (Corbin and Strauss 1990), which is a collection of techniques for both reducing data to comprehensible amounts, and then deriving useful theoretical insight from this summary data.

Grounded Theory has inspired a variety of derived methods and variants, collectively termed *Grounded Theory Methods*, or GTM (Braun and Clarke 2006). These methods excel at generating a strong explanatory narrative form collected data, and at seeing the unseen and relating the unrelated. GTM also allow for extracting multiple perspectives from the data

[8] http://redd.it/1m8vj6, active as at October 24 2014

[9] http://www.reddit.com/r/photoshopbattles/comments/1m8vj6/cat_killing_the_balloon/cc6va2j, active as at October 24 2014

(Boyatzis 1998, Braun and Clarke 2006).

When used in human-computer interaction research, GTM approaches typically follow a similar research programme (Muller and Kogan 2010): The research domain and type of data to be collected is established, and data collection is performed. Following the transcription of the collected data, the researcher spends time reviewing the data, and becoming intimately familiar with it. The codes, themes and categories from the data are then iteratively identified. These categories are then related to each other, and the conceptual structure of the data is identified.

It is important to note that research programme does not closely follow the original ideology of early Grounded Theory approaches, such as Glaser and Strauss (1967). However, this way of approaching the research fits well with the goals of HCI-focused research, as it allows for the development of the aforementioned explanatory understanding while still maintaining a specific initial focus.

One of the central tasks when undertaking a GTM approach in research is *coding*. Coding refers to the identification and extraction of key concepts and terms from the collected data, and several variants exist, each with a different specialty. *Open coding* is the first basic analytical step in GTM, and allows the researcher to begin conceptualising their data (Corbin and Strauss 1990). In open coding, concepts are identified and developed in terms of their dimensions and properties. These codes are then iterated: groups of codes are identified, and collectively summarised based on common thematic elements identified by the researcher. This iterative process then continues until the researcher has reduced the data to a volume from which they are able to derive a conceptual structure (Corbin and Strauss 1990, Dick 2005, Charmaz 2006, Star 2007). Grouping the data into categories and themes also allows the researcher to question the data with a

view to identifying new discoveries. The goal of GTM-based approaches is not detailed precision; rather, the researcher is more concerned about "recording any glimmer of themes or patterns" (Boyatzis 1998).

The broader term *grounded theory* describes a specific collection of analytical tasks that a researcher performs. However, as Corbin and Strauss (1990) notes, it is also useful and just as applicable to simply use individual components of GTM:

> *"Although if your purpose is just to pull out themes, then you could pretty much stop here [categories].* (Corbin and Strauss 1990)"

It is important to note that GTM are not "off-the-shelf" methodologies that provide specific instructions for researchers to follow. Rather, GTM are methods that provide guidance to researchers on how they should approach their data and make sense of it (Glaser and Strauss 1967, Muller and Kogan 2010). This means that GTM are unsuitable for the exploration of existing hypotheses (and nor are they designed for this purpose); rather, GTM are designed to enable researchers to *create* hypotheses from raw data (Suddaby 2006). In this sense, GTM are entirely inductive methods, rather than deductive.

Having completed both the semi-stuctured interviews discussed in Section 3.3 and selected candidate communities for analysis of their community guidelines, analysis of the collected data could commence.

An inductive approach to the data analysis, based on grounded theory, was developed and carried out for the analysis of this study. This approach to the analysis allowed for sufficient flexibility in the interpretation of the data, while still ensuring that findings derived from the data remained focused.

Traditionally, grounded theory-based methodologies seek to create an

explanatory discussion, or "theory", for the collected data. The advantages of grounded-theory methodologies include the fact that they allow the created theory to incorporate multiple perspectives on the same data, linking related parts together, and is useful in locating otherwise unseen information (Boyatzis 1998, Braun and Clarke 2006).

The approach used in the analysis of the data whose collection is reported upon in this chapter followed a fairly standard pattern for HCI research (Muller and Kogan 2010): the domain and type of the data to be collected was known, data was collated and codes, themes and categories were iteratively identified. While this approach does not precisely match the intended ideology of Glaser and Strauss (1967) and other early practitioners of grounded theory methods, the approach is an excellent fit for the goals of the project, due to it allowing for a thorough exploration of the collected data while still maintaining the core focus.

The GTM discussed in this chapter is based on an approach that merges methods proposed by Corbin and Strauss (1990), Dick (2005), Charmaz (2006), and Star (2007). As Glaser (1992) notes, the goal of the analysis is to produce new theory from the analysis, rather than evaluating or illustrating already-existing theoriues or ideas.

The components of the GTM-based appraoch were used to identify and extract key themes relating to differing types of manipulation on social media sites. The steps involved in this extraction, following Corbin and Strauss (1990), were:

1. *data familiarisation*: an initial review phase over all collected data, with researchers noting initial impressions and developing a strong familiarity with the source material;

2. *open coding*: identifying repeated phrases and words in the source

material.

To review: the data collected during interviews and from community guidelines was coded, and the most important codes were identified and grouped into themes. The most interesting themes were then selected. Following this, each selected theme is discussed (presented in Section 3.6).

The process is now presented in further detail, demonstrating how themes were identified and refined during analysis.

3.5.1 Data familiarisation

Data familiarisation is the first stage in any GTM. It is necessary for any researcher who aims to extract useful theory from raw data to be extremely comfortable and familiar with that data; to that end, researchers spend signficiant amounts of time simply reading and re-reading the information until they are satisifed that they are have a comfortable and intuitive understanding of the material. The purpose of data familiarisation is not memorisation; rather, the purpose is to give the researcher sufficient grounding from which they may begin the identification of codes.

The interview transcripts and subreddit rules were extensively reviewed multiple times before analysis commenced, until the researcher was extremely familiar with the content of the collected data. This took the form of repeatedly re-reading the interview transcripts, both in sections and in full, over the course of two weeks following the interviews; the subreddit rules were reviewed in the same manner.

Notes were collected during this processs regarding the author's thoughts on the interviews, but no coding took place during this period. When the data was felt to be sufficiently familiar to the author, the next phase of coding itself began.

3.5.2 Open codes

While grounded approaches frequently make use of multiple phases of increasing abstraction, Corbin and Strauss (1990) note that open codes on their own are themselves useful for the early exploration of a subject. To that end, open codes were identified and refined into the themes presented in this chapter.

The process of developing these codes involved reviewing both the transcripts of interviews, as well as the text of the community guidelines, and marking up important and repeated terms. The 'open' nature of the codes refers to the fact that, when coding began, no pre-defined codes existed. Rather, codes were identified during multiple passes through the text under analysis.

As an example, consider the following section from an interview with a moderator. The original spelling, capitalisation and corrections in the transcript have been preserved.

> "**Interviewer: Do false or spam submissions happen often?**
>
> Participant: *pics in aww are pretty fast judgement calls, same for WTF*
>
> Participant: *oh, constantly*
>
> Participant: *if you saw the stuff we don't let through, you might be amazed*
>
> **Interviewer: Can you talk about some examples?**
>
> Participant: *well, there's the easy-to-spotspot* [sic] *spammers who don't care if you know they are spamming, they just hope to get as many posts out there as they can before getting banned then there's the sneakier ones who might own multiple domains and try to evade detection that way*
>
> Participant: *they get creative, too*
>
> Participant: *once we're on to a domain, many times they start posting image links with the url superimposed on the image*

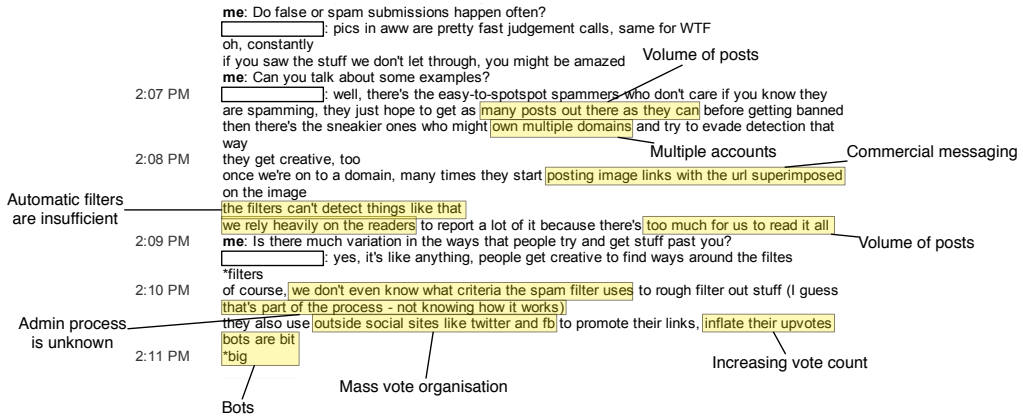


Figure 3.2: Example of the open coding process.

Participant: *the filters can't detect things like that*

Participant: *we rely heavily on the readers to report a lot of it because there's too much for us to read it all*

Interviewer: Is there much variation in the ways that people try and get stuff past you?

Participant: *yes, it's like anything, people get creative to find ways around the filtes* [sic]

Participant: **filters*

Participant: *of course, we don't even know what criteria the spam filter uses to rough filter out stuff (I guess that's part of the process - not knowing how it works)*

Participant: *they also use outside social sites like twitter and fb to promote their links, inflate their upvotes bots are bit* [sic]

Participant: **big"*

Figure 3.2 shows how this section of the interview was marked up, with initial codes determined.

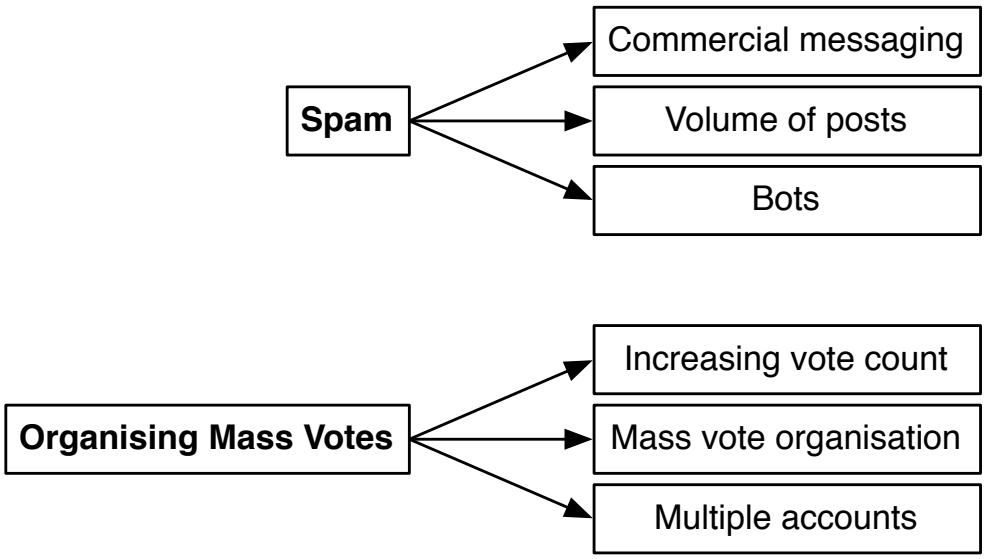


Figure 3.3: Example of refining codes.

3.5.3 Iteration of open codes

Having identified these initial codes, the codes were then grouped together by theme, focusing on types of manipulation. Codes that did not pertain to types of manipulation were not included in this process, but instead directly contributed to the definition of what manipulation is, discussed Section 3.7.

These grouped codes became themes, which form the basis of the discussion of manipulation on social media sites presented in this chapter.

3.6 Interpretation

The following themes were identified in both administrator interviews, and in subreddit analysis.

- Personality voting (*voting based on the identity of the poster, rather than*

the content)

- Spam (*Posting in order to directly sell services or products*)
- Attention grabbing (*Use of phrasing or typography in the post text to get reader attention, unrelated to the content of the post itself*)
- Rewarding upvotes (*Mentioning rewards or other benefits to the act of upvoting itself, such as promising a charity donation for each upvote*)
- Requesting upvotes (*Asking for upvotes without promising a reward*)
- Organising mass votes (*Asking groups of users to vote someone* else's *post either up or down*)
- Financial gain (*Posts that exist as part of a marketing campaign, either indirect or direct*)
- Post suppression (*Use of reddit's reporting and flagging tools to remove content*)

A distinction was made between "financial gain" and "spam" themes. "Spam" refers to direct marketing posts that urge people to buy goods and/or services, while "financial gain" posts are indirect, marketing posts.

Each of these themes are now discussed.

3.6.1 Personality voting

Personality voting is where users vote upwards or downwards based on the identity of the poster, rather than the content that they posted. If a user is well-known, participants reported that other users would up-vote them on the basis of their identity alone, and not of the content that was submitted:

> *"I think some of the high-karma 'power users' have probably figured out that if they get some people to upvote right at first, it helps."*

One particular subreddit mentioned by participants was */r/AskReddit*, which was described in Section 3.4.2. In particular, */r/AskReddit*'s first rule is:

> *"You must post a clear and direct question, and only the question, in your title. Any context, clarification, or conditions should be posted in the text box. Your own answer to the question should go in the comments as a reply to your own post.*
>
> *If you wish to tag your post NSFW, either put the letters 'NSFW' before or after the post, or use the tagger button. Introductory statements or claims, 'baiting' devices like 'Possibly NSFW', or non question-related information like 'I'll start...' are a violation of this rule, and will result in the post being removed."*

The reason for this change was described by a staff member at Reddit:

> *"They had a real problem with people using it as a platform for getting attention for themselves. For example, they'd post something like, "I just had sex for the first time! AskReddit, what was your first time like?" – you know, ostensibly posting the question to prompt a conversation, but really using it as an opportunity to talk about themselves."*

3.6.2 Spam

Spam, following Spamhaus Project's (2014) definition, is the bulk distribution of unsolicited messages. Spam is an issue that affects a broad spectrum of online communities, and is not limited to Reddit, or even to social media sites .

Participants reported that the administrators of the site - that is, employees of Reddit, Inc - have existing anti-spam policies that, while their existence is known, the details are kept confidential:

> *"Admins have the tools, which they won't reveal to us, even to /r/reportthespammers mods, we just have to play it by ear."*

Some communities explicitly approve of commercial postings. In most cases, these communities place limits on the number of self-promoting posts that an individual user may make. For example, the subreddit */r/ShutUpAndTakeMyMoney*, which focuses on novel products, specifies that users may promote their own goods, subject to limits:

> *"You may post your own products, but you MUST state that you are the creator and/or stand to benefit from the sale by either using the "Creator" flair or explicitly mentioning such in the title. In addition, you are allowed 1 post per month. (Shut Up and Take My Money moderators 2014)"*

One moderator reported that they spent significant amounts of their moderation time patrolling a subreddit named */r/reportthespammers*[10]. Unlike most subreddits, which are designed as a space for users of the site to discuss content, */r/reportthespammers* uses the existing reddit infrastructure to create a reporting system for general users: when a user believes that they have identified another user posting spam to reddit, they create a new post to */r/reportthespammers*, and link to the accused user.

An inductive finding discovered during interviews with moderators is that a distinction is made between "casual" spam and "professional"

[10] http://reddit.com/r/reportthespammers, active as at October 21 2013

spam. Moderators reported seeing submissions that were considered commercial in nature and therefore spam, but were not considered to be grounds for banning the user.

One of the interviewed moderators described a site operator who was accused of spamming Reddit:

> *"We talked, and the owner agreed to follow our rules - we would still allow him to submit his own site within those rules. We're both happy.*
>
> [Some] *are small-timers who simply don't realise they're spamming, from Reddit's point of view."*

3.6.3 Attention grabbing

Attention grabbing posts use manipulative phrasing and capitalisation to encourage users to click on them. For example, "I can't believe this just happened", or "OMG CLICK THIS!"

The subreddit rules for */r/worldnews* provide a good example of these kinds of restrictions:

> *"Do not editorialize the titles. No link shorteners / all caps / offensive / racist content. No editorial, opinion, petition, solicitation, poll or advocacy articles."*

An administrator described the use of attention-grabbing phrasing as "memetic manipulation" (referencing the concept of "memes" developed by Dawkins in 1989), in the following terms:

> *"It short circuits the content evaluation cycle of the reader."*

3.6.4 Rewarding upvotes

In order to receive upvotes, users may indicate that some benefit will be given, generally to charity, for each upvote received.

This behaviour is discouraged by the site-wide community guidelines, which apply to all subreddits hosted on the site (known as the "Reddiquette"), which states:

> *"Please do not ask for upvotes in exchange for gifts or prizes. "Upvote me to the top and I'll give away ..."*

Despite this ban, posts that fit this pattern still appear:

> *"Hurt me good r/atheism, $.50 to Doctors Without Borders for every upvote.*[11]*"*

3.6.5 Requesting upvotes

Users may explicitly ask for votes in the title of their submission. The specific phrasing varies from post to post, but generally either mentions a desire to make the post popular enough to be on the front-page of Reddit, a specific request for attention, or requests for upvotes as a result of some non-related reason.

An example of this form of manipulation occurred when Stephen Colbert, host of the US news entertainment show "The Colbert Report", mentioned on-air that he was a user of Reddit. Following this, a post with the following title was added:

[11] http://www.reddit.com/r/atheism/comments/myvu2/hurt_me_good_ratheism_50_to_doctors_without/, active as at October 24 2014

> *"Stephen Colbert just said he likes reading reddit. Lets get this on the frontpage and coerce him to do an interview/IAmA*[12]*"*

This post received 12,174 upvotes and 9,301 downvotes before being closed to new votes.

One frequently-cited reason for users to upvote a post is the user's "cakeday". Reddit displays a small image of a piece of birthday cake next to usernames on the anniversary of their joining the site, and users make reference to this anniversary as part of the post title. An example is this post, which is a photograph of two Siamese cats:

> *"In honor of my cake day, my boys are ready for your upvotes! They didn't get any love in r/aww. Maybe r/cats can appreciate them.*[13]*"*

The phrase "didn't get any love", seen in the above example, is a frequently visible manipulative element in submission titles, and is often seen attached to pictures of pets.

Another example of manipulative text in submission titles that actively request upvotes are "polling" posts: posts that encourage users to upvote or downvote based on some factor that is independent of the submission itself.

> *"Upvote if you wish clicking on the "special occasion" Reddit-logo-variation would take you to it's context. (à la Google)*[14]*"*

[12] http://www.reddit.com/r/reddit.com/comments/cksqt/stephen_colbert_just_said_he_likes_reading_reddit/, active as at October 24 2014

[13] http://www.reddit.com/r/cats/comments/ysnc4/in_honor_of_my_cake_day_my_boys_are_ready_for/, active as at October 24 2014

[14] http://www.reddit.com/r/reddit.com/comments/629m9/upvote_if_you_wish_clicking_on_the_special/, active as at October 24 2014

Participants reported that they attempted to prevent this kind of manipulation:

> *"We discourage such things; also, "cakeday" "vote up if", "dae"* ["does anyone else"] *etc, all manipulative."*

3.6.6 Organising mass votes

Any user on Reddit may upvote or downvote other posts, as long as the subreddit that the post is in is not private. The intended goal of the voting system, as has been discussed, is for users to vote up content that they find interesting and vote down content that they find disagreeable. Users are free to have other reasons, however.

Organising mass votes refers to a user organising to have a group of other users, either known to the user or not, vote upwards or downwards. This is disallowed by the site-wide community guidelines:

> *"Don't create mass downvote or upvote campaigns. This includes attacking a user's profile history when they say something bad and participating in karma party threads. (Reddit Inc. 2014a)"*

This message is reinforced by individual subreddits; for example, the community guidelines for the subreddit */r/ShitReddditSays*, which is discussed in more detail in Section 3.6.8, reinforces the site-wide prohibition on mass-voting in their own community guidelines:

> *"ShitRedditSays is not a downvote brigade. Do not downvote any comments in the threads linked from here! (Shit Reddit Says Moderators 2014)"*

A potential example of this kind of manipulation in action, which occurred some time after the data collection for this study completed, may be

found in a Reddit AMA ("ask me anything") interview conducted with the actor George Clooney[15]. During an AMA interview, an interview subject creates a post introducing themselves and offering to reply to questions; in this particular case, *"Hello reddit, George Clooney here. AMAA* [ask me almost anything]". Users may then reply to the post with their questions, and the interview subject posts follow-up comments to them to reply.

Users noticed that Clooney appeared to be replying to comments whose post score was negative - that is, the comments had more down-votes than they had up-votes. Users began to comment that Clooney was being extremely thoughtful, until a Reddit user proposed a potential alternative explanation:

> *"Dude, all these questions have negative karma. No idea why.*
>
> *Edit: Holy shit, I get it. Assholes ask questions, then go through and immediately downvote everyone else in the hopes their question gets answered. Which is a pretty shitty thing to do.*[16]*"*

In other words, according to this theory posited by the commenter in the above quote, questions were being replied to when they had a high ratio of upvotes to downvotes (and were therefore prominent and visible), but were later being downvoted by other users who wished to see *their* questions become more visible.

A moderator of the */r/IAmA* subreddit discussed the implications of this behaviour:

> *"This type of behaviour has become disturbingly common in AMAs,*

[15] http://redd.it/1wdzwq, active as at October 24 2014

[16] http://www.reddit.com/r/IAmA/comments/1wdzwq/hello_reddit_george_clooney_here_amaa/cf13x5i, active as at October 24 2014

and we have even considered not announcing the exact time of popular AMAs on the calendar so that users can't do this sort of thing.[17]"

3.6.7 Financial gain

The *financial gain* theme identifies posts and comments that pursue commercial goals, and are not direct marketing or sales pitches. This is differentiated from spam in that spam is direct marketing for products and services, while financial gain is indirect marketing.

3.6.8 Post suppression

Moderators have observed instances of users attempting to silence other commenters, through the use of publically-available history. One interviewed moderator described the use of "shaming bots", which "follow" users that have posted in controversial subreddits and attempt to draw attention to these activities.

> "[The bots] *will follow that user around and 'call them out', so to speak. Like, if someone posted on r/niggers* [a racist subreddit], *and then posted a comment on r/aww, a bot might reply 'You should know this user posts to r/niggers' to that person's comment in r/aww.*"

This form of manipulation is indirectly related to the phenomenon of organising mass voting campaigns (as described in Section 3.6.6), though is slightly more subtle: rather than explicitly inciting people to apply votes, this activity of alerting users to potentially objectionable previous posting history is apparently intended to encourage users to downvote comments and submissions posted by this user.

[17] http://www.reddit.com/r/IAmA/comments/1wdzwq/hello_reddit_george_clooney_here_amaa/cf13xqo, active as at October 24 2014

However, the success of this practice is not guaranteed: one moderator commented that:

> *"Sometimes the bot wins, and sometimes people find the bot annoying enough that they downvote it. Usually, people don't appreciate bots. It's additional noise in a place that's already loud enough."*

A bot that exhibits a related behaviour is *SRSTrackerBot*, an automomous script written by an anonymous Reddit user, which interacts with content on Reddit (Anonymous SRSTrackerBot author 2013). Bots are common on Reddit; perhaps the most prevalent is *AutoModerator*, a bot that searches for and automatically performs moderation actions (such as banning users, deleting posts, and so on) based on the content of posts.

The SRSTrackerBot continuously watches the top ten posts in the subreddit "Shit Reddit Says" (henceforth abbreviated as *SRS*.) SRS is a subreddit that identifies and discusses posts to Reddit that the subreddit's community finds *"bigoted, creepy, misogynistic, transphobic, racist, homophobic, or just reeking of unexamined, toxic privilege."* (Shit Reddit Says Moderators 2014). While the subreddit rules of SRS specifically forbid downvoting posts that are reported to the site, this is impossible for the subreddit moderators to enforce. The posts on SRS link directly to posts accused of being offensive or insensitive, which provides an easy method for readers of SRS to post comments and down-vote posts.

In response, the SRSTrackerBot watches the top posts in SRS and posts a warning in threads that appear, which follow the following pattern:

> *"Hello* [subreddit name],
>
> *This comment was submitted to /r/ShitRedditSays by* [user name] *and is trending as one of their top submissions.*
>
> *Please beware of trolling or any unusual downvote activity."*

Post suppression can also take the form of reporting users to administrators, who have the unconstrained ability to ban users and remove their comments from the site. One participant in the interviews presented in this survey recalled an incident in which a user was banned after apparent public outcry:

> *"It was shortly after some people had gone all vigilante trying to say that missing university kid was the Boston bomber.*[18] *We had gotten all kinds of messages that day from people complaining that the license plate was visible on the post.*
>
> *As we always do, we said license plates are considered public domain (they are); next thing I know, admin has not only removed the post, but banned the user (established account) for something people post on there all the time, no problem.*
>
> *I think there must have been a protest post about it somewhere, because we had too many complaints for it to be coming from the presence of the post itself.*
>
> *Definitely manipulation there - public pressure. It's akin to a type of 'bullying' from the readers."*

3.7 Discussion

This section discusses the implications of the findings noted in Section 3.6. To reiterate, the objectives of this chapter (previously presented in Section 3.2.1 were to first establish whether administrators and moderators

[18] In March 2013, two bombs detonated at the finishing line of the Boston Marathon, killing 3 people and wounding an estimated 264 others (Kotz 2013). In response, a subreddit community named */r/findbostonbombers* was created, in which users attempted to analyse photos and other evidence in an effort to identify the perpetrators. The subreddit mistakenly identified a university student, who had gone missing one month prior to the bombing. The student was later found dead (Stanglin 2013).

consider manipulation to be present, and then to identify the different types of manipulation that they observed.

3.7.1 Manipulation exists

The interviews conducted during this study conclusively indicate that all participants believe ranking manipulation to be present in the social media site that they moderate.

This conclusion is evident in the analysis of subreddit rules and in conversation with administrators and moderators, and may be reached through independent analysis of each: both subreddit rules and moderator interviews reveal this information, though in different ways.

3.7.1.1 Manipulation is evidenced by subreddit rules

Subreddit rules reveal the existence of manipulation through the prohibition of specific types of content.

A common element alluded to in Section 3.4.1 was the fact that multiple subreddits prohibit the posting of "image macros", or images with superimposed text. An example of this type of content is shown in Figure 3.4.

Certain subreddits forbid not just images that follow the same format as Figure 3.4, but also the entire practice of posting content with the express purpose of gaining votes for oneself, as opposed to providing interesting content to the community. The rules for the subreddit */r/GameOfThrones* contain an explicit "What Not To Post" section, in which they provide concrete examples of material they consider to be contrary to the interests of the community's members; in one particularly striking example, the subreddit rules state that while a link to a store that sells replica swords from the television show is permitted, a "*photo of you holding* [the]

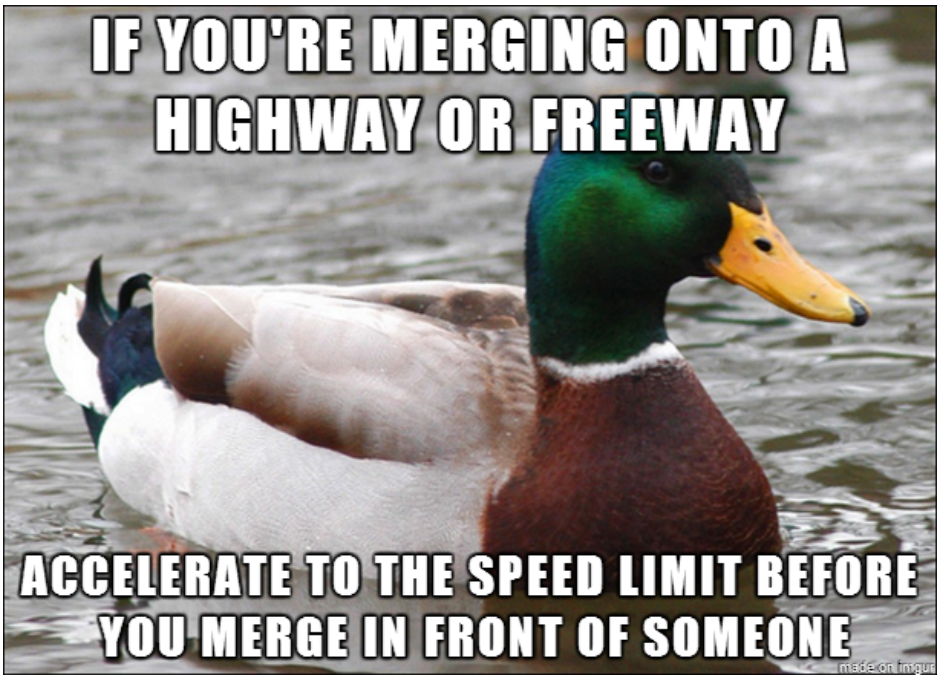


Figure 3.4: An example of an 'image macro'. The photograph of a duck is a stock photograph, and is known as "Advice Mallard"; image macros that use this photograph typically present what the author either genuinely or satirically believes is good advice. This image was taken from the front page of */r/AdviceAnimals*, a community for sharing similar images.

sword for karma" is not - casting this in the terms established in Section 3.6, the "reference to self" form of manipulation is expressly forbidden.

It is therefore reasonable to infer that the reason such content is forbidden is that, prior to the establishment of the rule prohibiting their posting, users of these communities were seeing too many of these kinds of images.

On Reddit, the only way for any content to be visible to large numbers of users is for it to receive up-votes, and consequently have a high enough ranking for it to appear on the front page. Consequently, a sufficient number of image macro posts had to have received a high number of votes in order to appear; however moderators clearly did not feel that this form of content did not fit with their opinions of the types of content that the community should promote.

At this point, it is important to highlight the fact that administrators and moderators comprise a different population of users than non-moderating members of the community. Moderators and users may disagree on what constitutes appropriate content for a community, though only moderators have the ability to enforce their decisions on content in the community (through the use of tools such as banning users, deleting or editing posts, and other administrative tasks).

A user's only recourse, if they disagree with the policies of a community's moderators, is to leave the community and join or create a new one. Examples of this include the subreddit */r/pics*, which has very strict and specific prohibitions on certain kinds of images: as a reaction against these rules, so-called "anything-goes" communities exist, with much fewer rules, such as */r/AnythingGoesPics*[19], *active as at October 24 2014* and */r/images*[20], *active as at October 24 2014*.

[19] http://reddit.com/r/AnythingGoesPics

[20] http://reddit.com/r/images

However, the subreddit rules of these anti-prohibition subreddits themselves reveal the existence of moderator-perceived manipulation: the subreddit rules of */r/AnythingGoesPics* (a subreddit whose rules were not included in the set of communities analysed in this study) prohibit *"solicitations"*:

> *"No Solicitations allowed. This isn't Craigslist. Tagging Post with NSFW tag on submissions with such Content is Required."*

The context of the word "solicitations" is explained by the phrase *"this isn't Craigslist"*: the online classifieds site Craiglist (Craigslist 2014) has a reputation for being a site where encounters with prosititutes may be arranged (Lambert 2007). The reader may therefore reasonably infer that a type of content that the moderators of this self-declared "anything-goes" community have seen an influx of prostitute solicitations, which they consider to be a form of posting that they do not wish to see in the community that they moderate (the potentially illegal nature of online solicitation by sex workers notwithstanding.)

If moderators find a community overwhelmed with a certain kind of content, and feel that its prevalence is counter to the community goals, the establishment of subreddit rules that prohibit its content is evidence that they believe that type of content receives more votes than is due, and the regular voting system provided by Reddit's infrastructure is insufficient for dealing with this type of content.

Consequently, subreddit rules indicate that ranking manipulation exists.

3.7.1.2 Manipulation is evidenced through interviews with administrators and moderators

Interviews with moderators and administrators indicate that they believe that manipulation exists, and is a problem in the communities that they moderate.

All moderators, when asked whether they thought manipulation was a problem, emphatically agreed. Administrators of the site, who operate at a higher level of moderation than the moderators of individual communities, agreed with the statement that manipulation is a phenomenon on Reddit.

However, the degree to which they thought it was a problem varied between different types, and between moderators: one participant, who described himself as highly spam-focused, noted the difference between "casual" and "professional" manipulation, and noted that he only cared significantly about professional manipulation.

3.7.2 Manipulation attempts to influence relevance

Social media sites are environments in which users may both post content and vote on how prominently content should be ranked. As a result, users post content that they hope becomes highly-ranked. The public nature of how users vote means that submitted content will attain a certain ranking, depending on several factors including the thoughts of the first users to see the content, the time and date at which the content is posted, and the specific community in which the content is posted.

Because content on social media sites is listed linearly, it is not possible for two pieces of content to share an equal degree of prominence. Accordingly, all content that is ranked highly is ranked at the expense of other,

lower-ranked content. Any attempt to improve the ranking, therefore, is done at the expense of other content.

As has been shown in this chapter, manipulation behaviour takes several forms:

- Users can post obvious **spam**, or original content that is clearly intended for **financial gain**.
- Users can use **attention-grabbing** techniques to focus user attention on their content, which increases the chance of users voting the content up.
- Users can arrange for additional votes to be cast towards their content, either through **directly paying for votes**, **indirectly rewarding users for their votes** or simply directly **asking for votes**.
- Users that are **well-known personalities** have a tendency, according to moderators, to receive more votes than others, regardless of the content they post.
- Users can arrange to have down-votes cast on content that they dislike, such as by **organising mass-voting campaigns**, or use techniques to **suppress posts** they dislike, like incorrectly reporting content as spam, in order to reduce the prominence of other content.

In the context of the work on relevance by Saracevic (2007) (and previously discussed in Section 2.3.7), these forms of behaviour are attempts to increase the relevance of content.

To review: Saracevic, in his 1975 paper, generalised definitions of relevance into a pattern that described relevance as the measure by which information is applicable as its is conducted from one agent to another.

Following this, Saracevic's later work (in 2007) built upon this generalised definition and identified a multiplicity of different types of relevance. Among these definitions, we find:

- *Topical* relevance, which refers to the *aboutness* of the information;
- *Affective* relevance, which refers to the *satisfaction* of the information seeker; and
- *Algorithmic* relevance, which refers to the *relative effectiveness* of content as decided by a sorting algorithm[21].

In the context of social media sites, all users seek content for a variety of reasons. These reasons are effectively represented by the types of relevance identified in Saracevic's work: users visit a social media site looking for topical information, such as in topic-focused communities like the *"Game of Thrones"* subreddit, or looking for "satisfying" information, such as the *"WTF"* subreddit[22].

However, the specific content offered to them by the Reddit software employs an additional type of relevance: algorithmic relevance. The Reddit software uses the age of a piece of content together with the total positive and negative votes that that content has received in order to determine its relative ordering (Salihefendic 2010). When users find relevant content that they approve of, they upvote it. This voting has a direct influence on the algorithmic relevance of that content.

[21] Borlund (2003) notes that algorithmic relevance frequently does not operate in isolation; for example, topical relevance is often used as part of the design of these algorithms that determine relevance.

[22] The term 'satisfying' here is used in the sense of satisfying the user's desire to see surprising content; much of the content posted to the *"WTF"* subreddit is decidedly *not* satisfying in the sense of it being pleasant to view.

Revisiting the forms of manipulative behaviour listed above, we can now re-contextualise the manipulative behaviour in terms of the types of relevance that are at play:

1. Attention-grabbing techniques attempt to increase the topical and affective relevance of content, through editorialised titles and use of capitalisation.

2. Payment for either up-votes or down-votes on content directly influence the algorithmic relevance of content.

3. The presence of well-known users influences the affective relevance of content by associating their identity with the content.

We must now consider the fact that the topical and affective relevance of content are specific to individual users, while algorithmic relevance applies to all users. As Saracevic (2007) notes, *topical* relevance refers to whether the topic of the information matches the topic that the reader is seeking, and *affective* relevance refers to the individual satisfaction of the reader. Both of these types of relevance pertain directly to the state of the reader as an individual, and can therefore be considered, within the context of social media sites, *individual*-focused types of relevance.

However, the impact of *algorithmic* relevance is not limited to individual users. Social media sites use votes provided by their users to determine the relative ordering of content; this ordering applies to all users, and does not generally vary significantly between users[23]. Algorithmic relevance, in the context of social media sites, is therefore considered a *global* property.

[23]Reddit allows users to customise their feeds, but only insofar as which subreddits they choose to display on their home page. The order of the subreddit content remains dependent on the content's vote tally and its age.

Having established the domains of influence of these types of relevance in social media sites, it is now possible to establish a definition of manipulation itself. This chapter has shown that different kinds of relevance are at play when users interact with a social media site, and that both topical and affective relevance influence of content affect users on an individual scale, while algorithmic relevance of content affects all users of the site.

When users find content that they find topically or affectively relevant, they vote that content up. Because the tally of votes is used by the site's ranking algorithm to determine relative content prominence, when a user votes up content, it becomes slightly more prominent for all users. The cumulative effect of votes determines the prominence of all content.

When a manipulative technique is employed, the content poster seeks to influence individual readers. As has already been discussed, the methods involved include attention-grabbing techniques, directly asking for votes, and so on. These are attempts to influence the topical and affective relevance of the content. In doing so, content posters who use these techniques are attempting to indirectly influence the algorithmic relevance of the same content.

It is important to note that manipulation does not involve any *direct* attempt to modify algorithmic relevance, and instead relies on indirect methods. Users who wish to promote their content and who do not have direct control over the site itself cannot directly modify the ranking algorithm that the site uses, and must must play by the rules that this algorithm imposes upon all users. This means that a site administrator or community moderator - someone who possesses control over the site that users do not - could indeed promote their content over other users; this is a common occurrence, and is usually referred to as "pinning" content, so as to ensure that all visitors to the site can see it. However, because this direct

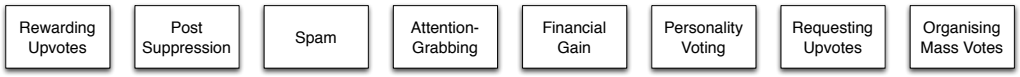


Figure 3.5: The initial framework of manipulation, presenting the types of manipulation identified in Phase 1.

approach does not involve any techniques that require participation from non-privileged users, it is not manipulation.

A definition of manipulation now becomes possible: *manipulation on social media sites is an attempt to change the global algorithmic relevance of content by changing its individual topical and affective relevance*.

This definition is used throughout the remainder of this thesis as a focus for the discussion of manipulation on social media sites.

3.8 Summary

This chapter has reported on semi-structured interviews that took place with moderators and administrators at Reddit, in conjunction with the analysis of 16 subreddit community guidelines. This resulted in the classification of different types of manipulation, which is put to use in the later phases of the research, discussed in Chapter 4 and Chapter 5.

Broadly, this study has fulfilled the objectives for which it was designed, which are discussed in Section 3.2.1, in the following ways:

- Manipulation has been established as an issue for moderators and administrators. This validates the underlying proposition of this research (namely, that manipulation exists, and can be studied).
- A definition of manipulation has been established, which permits the following phases of this research (presented in Chapter 4 and Chap-

ter 5) to follow a more constrained focus. This definition is framed in the context of Saracevic's (1975, 2007) work on relevance.

- A number of types of number of types of manipulation have been identified and classified, forming the basis for tools with which to study manipulation in social media sites.

Therefore, it is argued that this chapter has been successful in meeting its objectives. A wide range of findings has been presented, and the information gathered from these serves to validate the research premise.

4

Phase 2: Diary study

This chapter reports on a web-based study, in which a web browser extension was developed to collect data from participants in order to explore types of manipulation identified by interviews and in the community guidelines from subreddits on Reddit, as discussed in Chapter 3.

As a result, the categories of manipulation previously identified by administrators, which was discussed in Chapter 3, are further developed and elaborated. This provides a more complete context in which to discuss manipulation within social media sites, and is instrumental in the construction of a model of manipulation, presented in Chapter 6.

4.1 Introduction

Having established guidelines for classifying forms of manipulation from administrators and moderators in Section 3.6, it was then necessary to determine how frequently users of Reddit observed manipulation, and to develop an understanding of user perspectives on manipulation.

A diary study was implemented that collected information directly from users. This allowed the study to gather immediate, up-to-the-minute

data without forcing participants to leave the context from which data was being collected. This diary study was carried out using a novel in-situ method, involving a web browser extension that allowed users to participate in the study while browsing the site.

This study resulted in the further exploration of each of the previously identified types of manipulation discussed in Chapter 3. In Chapter 3, administrators and moderators of a social media site were interviewed, from which the transcripts were analysed alongside the text of community guidelines in place, which were also written by moderators. This resulted in the development of a classification of manipulation types, but was limited in its perspective to that of administrators and moderators. In order to evaluate the completeness of this classification, and to determine how , it was necessary to gather perspectives from users of the site.

This chapter discusses the design, implementation and deployment of the diary study, and presents an analysis of the data collected. This phase found that the most prevalent form of manipulation was attention-grabbing, which has multiple facets. Additionally, it was found that participants may have a non-standard definition of what spam is.

4.1.1 Chapter structure

This structure of this chapter is as follows:

- Section 4.2 provides an overview of the study, and the objectives served by it. Ethical considerations are discussed, and the contributions provided to the overall research are enumerated.
- Section 4.3 presents the methodology and design of the study and discusses the development and deployment of the research tool created to gather the data reported upon in this chapter.

- Section 4.4 reviews the findings of the study, and presents an elaboration of the types of manipulation identified in Chapter 3.
- Section 4.5 discusses the implications of the findings presented in the previous section upon the research as a whole, and presents the contributions of the chapter.
- Section 4.6 concludes the chapter, and discusses the next steps for the research.

4.2 Approach

Phase 2 of the research was intended to gather user perspectives on manipulation in social media sites, and to explore the framework created in Chapter 3.

In the process of carrying out this intent, this phase of the research also served to evaluate the use of an embedded real-time data collection methodology, which served to quickly and efficiently gather in-situ data from study participants.

Finally, the analysis of this study allowed for further refinement of the analysis whose findings were reported upon in Section 3.6. By exposing the categories derived from the research conduted in Chapter 3, participants were able to provide focused data that was used to elaborate on the categories and allow for further exploration of the topic.

This section is structured as follows:

- Section 4.2.1 discusses the objectives of the study described in this chapter.
- Section 4.2.2 discusses the ethical considerations taken into account in the design of this research.

- Section 4.2.3 discusses the contributions made by this chapter.
- Section 4.2.4 discusses the method and approach taken to gather and analyse the data.

4.2.1 Objectives

The objectives of this study were:

1. *to explore the relative prevalence of the various types of manipulation identified in Chapter 3, and to gather user perspectives on what they considered to be manipulation*.

   This served to both verify that the types of manipulation identified were valid, and to further explore the various manifestations of those types of manipulation.

2. *to gather information on user perspectives on manipulation, using the framework of the classified manipulation types discussed in Chapter 3*;
3. *to undertake a second-round analysis of the data collected in the first phase, supported by additional data collection*; and
4. *to evaluate the use of a web browser extension to collect in-situ data from participants browsing a web site*.

These chapter-specific objectives serve the overall objectives of the thesis by providing additional perspectives on manipulation.

4.2.2 Ethics

The study described in this section was approved as a Minimal Risk Study by the Tasmanian Human Research Ethics Committee. The reference number for this study was H0013123.

To ensure the privacy of the participants and to reduce bias, the following steps were taken:

- Participants were not required, at any time, to provide any personal information.
- Participants were informed that information that they submitted as part of the study was not visible to any entities
- Participants were free to withdraw from the study at any time.
- While user IP addresses were collected in order to facilitate handling of potential abuse, this information was discarded prior to analysis.
- Likewise, while the user's (pseudonymous) username was collected, this information was only used for grouping data, and was discarded prior to analysis.

4.2.3 Contributions

This chapter makes the following contributions to the overall research:

- A refinement of the framework of manipulation and its various impacts upon an online community
- The evaluation of a web-browser extension designed to collect specific, in-situ data from participants.

These contributions assist the research in addressing the research objectives stated in Section 1.2. Specifically:

- By refining the framework that describes manipulation, the research is better able to address RQ1, which asks, *What are the most prevalent kinds of manipulations taking place on these social media sites?*.

- By collecting information in-situ from users of a social media site, the research able to gather direct data to address RQ2, which asks *What impact do these types of manipulations have on the communities?*

4.2.4 Design

In order to address the first objective of this chapter, information needed to be gathered from users. The number of users on Reddit greatly outnumbers the moderators and administrators. This demanded a novel approach for collecting data from this class of participant, which emphasised the ease of recruitment, minimal reporting burden placed upon each participant, and immediate data collection in order to prevent data loss from delinquent participants (that is, users who dropped out of the study).

In addition to these considerations required by the non-exclusive nature of participant recruitment, there were further methodological reasons for immediate data collection. As suggested by Czerwinski, Horvitz and Wilhite (2004), diary studies in which participants log data "after-the-fact" mean that there is a delay between a reportable event and the participant noting it in the diary. This can lead to inaccuracies in the logged data; additionally, any delay between an event taking place and it being logged increases the chance that the information is never logged in the first place.

Diary studies can be burdensome for their participants. Rieman (1993) presents an example in which users were asked to log their behaviour for a period of a week, and found that the logging requirements proved too onerous for some participants, and reduced the quality of the data they collected. Accordingly, the diary study discussed in this chapter was designed to be as minimally intrusive as possible, and to require as little effort from the user as possible.

Finally, by collecting data immediately, the researcher was able to re-

ceive information at the moment of it being recorded. This prevented cases of users never submitting their collected information to the study: if a participant became bored with the study after a time, all information gathered from them up to the point where they left the study had already been collected in its entirety.

In taking these constraints into account, the diary study was implemented in a novel way, by using a browser extension to embed the study directly into the user's experience of the social media website. The advantages of using a web browser extension for data collection in this study were numerous. Because an extension is situated in the browser, which the user is already using, participants do not have to suspend their browsing activity in order to record information, which have been noted in other diary studies to disrupt the activity under observation (Czerwinski et al. 2004). This allows for results to be reported at the same time and in the same location as the observation.

Diary studies are discussed in more detail in Section 4.3.2; additionally, academic and industry uses of web browser extensions are discussed in Section 4.3.2.1.

4.3 Online data collection

The data in this study was collected through the use of a web browser extension, installed by each participant, which allowed them to report when they believed they were seeing an example of manipulation on the site Reddit.com.

The structure of this discussion of the data collection is as follows:

- Section 4.3.1 discusses the scope of the study, and the various constraints elected to be used by the researchers.

- Section 4.3.2 discusses the choice of methodology used in this study, and provides further information on how the construction, testing and deployment of the data collection tool was carried out.
- Section 4.3.3 presents the design of the study itself, and provides more detailed information on how the data collection software used by participants operated.

4.3.1 Scope

The scope of this study was quite straightforward: to gather examples of manipulative content, classified by users according to the classification scheme proposed in Chapter 3, as well as gather optional comments.

Due to the characteristics of the web-based data collection tool, the scope constraints placed upon the study were primarily technical in nature. At the time of the web browser extension's development, the four most popular web browsers in use were Microsoft's *Internet Explorer* (Microsoft Corp. 2014), Google's *Chrome* (Google Inc. 2014), Mozilla Corporation's *Firefox* (Mozilla Corp. 2014), and Apple's *Safari* (Apple Inc. 2014).

The web browser extension was initially developed for Safari, due to the investigator's experience in developing for this browser. Once development was complete, the software was then adapted for use in Google's Chrome browser. At the time of development (October 2013), Chrome and Safari together controlled approximately 46% of global desktop web browser usage (StatCounter Inc. 2013).

4.3.2 Choice of methodology

The methodology used in this study was adapted from the practice of diary studies. Typically, diary studies involve participants recording their

personal experience in a form that is later collected by the researcher for analysis; in the case of this study, a more immediate form of data collection was desirable. Therefore, the design of the study took into account a requirement that all data collected by participants must be immediately transmitted to the researcher, rather than being stored for later collection by the researcher. This is referred to as *real-time* data collection.

The primary reason for the requirement that data collection occur in real-time was that all participants were recruited anonymously over the internet. This posed a potential challenge for collecting data from users, due to the fact that follow-up was more difficult; the software was able to be downloaded by any individual, which meant that the identity of each user could not be ascertained. (The implications of this fact also have effects on how the third phase of the research, as discussed in Chapter 5, was carried out.) To that end, immediate data collection proved invaluable to the operation of the study, as it ensured that all data recorded by users was immediately made available for analysis.

The second reason for live data collection was that it enabled the researchers to identify and correct certain kinds issues in the software, without having to wait for bug reports. Prior to the public release of the data collection tool, important bugs were identified by the author based entirely on the initial data that was received during the initial pilot testing; the data collection tool was then able to be updated, and automically downloaded to all existing participants.

In addition to the problems already noted in this section, diary studies also suffer from the problem of split attention (Czerwinski et al. 2004). Typically, diary studies record their data in a different context from the source of the data itself; for example, a diary study concerning the number of times a person sees a duck while jogging must involve users requires

the user to stop jogging for a moment while writing down their duck experience. As Czerwinski et al. (2004) notes, this can create a potential loss of information in between the moment of observation and the moment of recording of that information.

To address this issue, the diary study discussed in this chapter used a method of embedding the study itself in the user's web browser, which allowed for in-situ data collection.

4.3.2.1 Web browser extensions

A web browser extension is a piece of software that runs in the context of a web browser. Web browser extensions are capable of extending the user interface of the web browser, inspecting the contents of the web pages that the user is viewing through the web browser, and injecting content into these web pages.

One of the most popular examples of web browser extensions are advertisement blocking extensions. These extensions maintain a list of sites known to serve online advertisements from, and modify the content of each web site viewed through the browser to remove any images or Flash animations served from the list, which has the effect of hiding advertisements (Singh and Potdar 2009).

In addition to removing content, web browser extensions are able to insert new content into pages. A popular example of this is the Reddit Enhancement Suite, which inserts additional functionality into the Reddit website (Sobel 2014).

The key benefit for web browser extensions that provide additional content is that the user is able to view this additional content without having to stop viewing the website. This advantage was critical to the implementation of the diary study.

Finally, the experience of the author in the field of software development played an important role. As a developer of mobile and web-based software since the mid-2000's and the co-author of several works on software development (Goldstein, Manning and Buttfield-Addison 2010, Buttfield-Addison, Manning and Nugent 2014, Manning and Buttfield-Addison 2014), the author's skills in the rapid development of software tools meant that the necessary functionality could be rapidly developed and deployed.

4.3.2.2 Development of the web browser extension

The data collection tool was comprised of two components: the web browser extension itself, which ran on individual participants' computers, and a server-side component that received data from the web browser extension. For brevity, the web browser extension is referred to as the *client*, and the server-side componet is referred to as the *server*.

The development of the data collection tool was divided into three phases: *initial development*, in which prototypes of the client and server were created; *pilot testing*, in which the prototypes were deployed and tested with an initial, limited group of people, resulting in improvements to the prototypes, and *public deployment*, in which the client was released to the public, and data collection began.

Initial Development The web browser extension comprised two components: a front-end component written in JavaScript (European Computer Manufacturers Association 2011), and a back-end component written in Python (Python Software Foundation 2001). The extension was originally written for Apple's Safari web browser, Google Chrome and Opera (Opera Software ASA 2014). The back-end component of the web browser ex-

tension was written using the Flask framework (Grinberg 2014), and was hosted on Heroku (Heroku Inc. 2014), a platform-as-a-service product that allowed for rapidly iterative development.

The decisions for these choices of technology were straightforward: on the majority of popular web browsers (namely, Google Chrome, Firefox, Opera and Safari), JavaScript is the only option for developing web browser extensions. Python is well-regarded as an effective language for server-side web application development (Grinberg 2014), and the Flask framework provides an effective and well-tested implementation of many common functions needed by web-based applications, which simplified development and reduced the opportunity for bugs and defects in the application.

The choice of Heroku reflects a developing trend in web-based application development, in which web applications are not hosted on dedicated computers owned and maintained by the web application developer, but instead are hosted on machines owned and operated by a third party, who manages considerations such as system administration (Armbrust, Fox, Griffith, Joseph, Katz, Konwinski, Lee, Patterson, Rabkin and Stoica 2009). This trend, frequently referred to as "cloud computing" (Mell and Grance 2011), allows developers to focus on the application-specific logic needed by their product, rather than the day-to-day operation of the systems that enable the application to function.

In addition to receiving information from the distributed installations of the client software, the server component also provided the mechanism that allowed for remote updating of the client software.

Pilot Testing During the pilot testing phase, the client was distributed to a small group of people known to the developer. The author also posted on

Twitter, seeking additional volunteers who wished to participate in early testing of the software. An initial testing group of 10 volunteers downloaded the software and began logging data.

For a period of two weeks, testing of the client was undertaken. Testers were asked to use the data collection tool to begin gathering examples of what they believed to be examples of manipulation on Reddit, and to report any problems with the client. During the testing phase, the author reviewed incoming data to ensure that the data collection tool was reporting accurate and useful information.

This pilot phase proved invaluable in improving the usability and usefulness of the data collection tool. For example several bugs were identified during testing that caused websites other than Reddit to display the manipulation interface; another issue discovered during the pilot testing phase included an issue in which comments on Reddit (that is, user comments attached to posts) did not include any means of permitting the researcher to read the content of the comment being reported. Due to the fact that the web browser extension was able to be remotely updated, these were able to be corrected within 24 hours of being reported.

After two weeks of private testing, the data collection tool was released to the public for general participation.

4.3.3 Design

The web browser extension used in this study was designed to be as minimally intrusive as possible, in order for the data recorded by participants to be as representative as possible of normal use of the site. Accordingly, the design goals of the software were:

1. The software must not modify or hide any content normally pre-

sented by the site: each participant must view the same content as they would had they not installed the software.

2. The software must be entirely unambiguous about what content they are reporting as manipulation.

3. The software must impose as minimal a burden on the user's time and attention as possible.

In order to satisfy these design goals, the design of the software was as follows.

When installed, all posts and comments that the user saw on Reddit included a small link titled "manipulation?", as shown in Figure 4.1. This link was included alongside the pre-existing management links that the Reddit site already includes, which allow the user to flag content as inappropriate, reply to the user who posted it, and similar tasks. By including the "manipulation?" link in this area, the "look and feel" of the site was preserved, and the software did not introduce any additional visual burden on the user.

This inserted link was the only visible element visible to the user until they took any kind of action. No content was hidden, modified or inserted, which preserved the browsing experience of the site. Additionally, by including a "manipulation?" link under every comment and post, the association between the link inserted by the software and the content that would be reported upon by clicking the link was made unabmiguous.

When the "manipulation?" link was clicked, a dialog box appeared, as shown in Figure 4.2.

The dialog box allowed the user to select which type of manipulation they believed they were seeing. The choices available were derived from

Figure 4.1: The "manipulation?" link is appended to the already existing list of links present underneath every post.

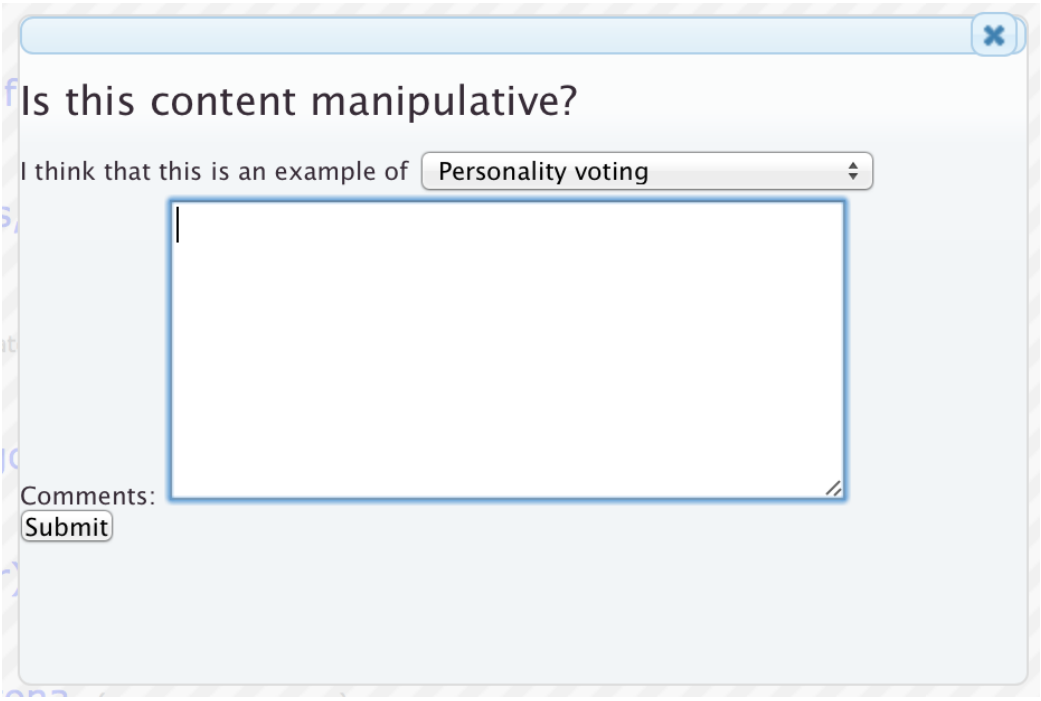


Figure 4.2: The dialog box that appeared when a "manipulation?" link was clicked.

the types of manipulation identified by administrators in Section 3.6: attention grabbing, financial gain, organising mass votes, personality voting, post suppression, requesting upvotes, rewarding upvotes, and spam.

Additionally, participants could optionally provide additional free-form commentary in a text field, which allowed users to provide any explanations they felt were merited when reporting content.

When the participant clicked the "Submit" button, the dialog box immediately disappeared, and returned the user to browsing Reddit. Simultaneously, the web browser transmitted the following data to a server controlled by the researchers:

- The selected type of manipulation
- The contents of the comments text field
- The item ID of the comment or post that had been flagged
- The first several characters of the user ID of the currently logged-in user, if the user was signed in

When the server received this information, it was immediately stored in a PostgreSQL (PostgreSQL Global Development Group 1996) database. The researcher was then able to download the entire set of data received. An 18-character randomly-generated password was used to prevent unauthorised access to the information.

In addition to the characteristics noted above, the server also recorded the time and date of the submission, as well as the IP address of the computer sending the information. While the user ID and IP address of each participant were recorded, this information was used only to mitigate against abuse (such as multiple submissions from the same user or the same IP ad-

dress), and the information was discarded prior to analysis. No personal information was used as part of data analysis.

By transmitting the data to the server in the background and returning the user to browsing immediately, the software significantly reduced the burden placed upon the user: instead of waiting for the data to be transmitted before they could return to browsing, the process became significantly more seamless.

However, this approach is not without its drawbacks: by returning the user to browsing immediately, the software loses an opportunity to inform the user if there is a problem in transmitting the data to the server. This was viewed as an acceptable risk by the investigators, as it was judged more important to reduce the burden on the participants than it was to create a more complex system for repeated attempts. The researchers' justification for this implementation is that individual data points are not critical information on their own, and are only useful in the aggregate. Therefore, while it is probably that a percentage of collected was lost in transmission, this does not significantly impact the data collection as a whole.

4.3.4 Participation

An announcement post was made on Reddit, providing a link to information about the study and an open invitation to begin participating. Participants were informed of the background of the study and given examples of manipulation based on the preliminary insights derived from the work performed in Chapter 3; specifically, *attention-grabbing*, *requesting upvotes*, and *organising mass votes*.

The announcement was posted twice, once each to the subreddits */r/The-*

oryOfReddit[1] and */r/HailCorporate*[2].

Ironically, the posting of the announcement initially fell victim to a measure designed to counter manipulation. When the author emailed one of the Reddit site administrators and mentioned that the second phase had begun, the administrator replied saying that he had noticed that the announcements had not resulted in any significant uptake; upon looking into it, he emailed the author, saying that he had discovered that the IP address from which the announcements had been posted was *shadow-banned*.

Shadow-banning, also known as *hell-banning* (Lewis 2014) is a technique[3] for countering abusive users, in which any comments or posts created by the user only appear for that user, and are hidden from all other users. This prevents or delays the user's realisation that they have been banned; once they realise their situation, they may disguise their identity, re-join the community, and resume posting content. Shadow-bans are similar to Usenet's *kill files*, which predates shadow-bans: a kill file lists strings of text that the user does not want to see, and which the user's Usenet client automatically filters out (Smith and Kollock 1999). Because kill files do not prevent users from posting content (and instead prevent other users from seeing it), they will remain unaware that other users are not seeing their content. Shadow-bans, and related techniques, are discussed further in Section 6.3.2.5.

The shadow-ban applied to the author's IP address was graciously removed by the author's administrator contact, and the announcement was

[1] http://www.reddit.com/r/TheoryOfReddit/comments/1p36dw/im_running_a_study_on_how_people_manipulate/, active as at October 24 2014

[2] http://www.reddit.com/r/HailCorporate/comments/1ot25e/im_running_a_study_on_how_people_manipulate/, active as at October 24 2014

[3] Lewis (2014) describes shadow-banning as a *temporal dark pattern*: a reusable design technique for wasting a user's time.

re-posted; the reason for the shadow-banning was apparently, according to the administrator, due to a tendency for the author's internet service provider to change IP addresses rapidly, which was mistaken for an attempt to circumvent anti-spamming measures.

Once the announcement had been successfully posted, the announcement posts reached the front page of the subreddits they had been posted to; on the subreddit */r/TheoryOfReddit*, the extension was vigorously discussed, with over 150 comments attached.

Approximately 30 individuals participated in the study; it is impossible to determine the precise figure, due to the anonymous nature of the study, but it is possible to develop an estimate by reviewing the user ID fragments that were submitted. Several hundred pieces of potential manipulation were reported by users of the diary study extension during the study period; these were either links to posts on Reddit, or links to the text of comments on Reddit. In the case of posts, the title is shown, while in the case of comments, the text of the comment is shown.

No material incentive, financial or otherwise, was offered in return for participation in the study. Participants who had taken part in Phase 1 were not explicitly recruited; however, the anonymous nature of participation means that previous participants could not be excluded.

4.3.5 Analysis

The analysis of the data collected during the study was conducted through a grounded-theory influenced approach. The reader is referred to Section 3.5 for a more detailed discussion of the grounded-theory-based approach used in the analysis of this study.

Unlike the data collected in Chapter 3, the data collected in this study did not take the form of interview transcripts; rather, the information col-

lected primarily took the form of URLs that linked to posts and comments on Reddit, which were associated with different types of manipulation as described in Section 3.6. Analysing this information, therefore, involved visiting each reported link, and analysing the content found there. In a strict sense, the data submitted by participants was not the raw data that was directly analysed; rather, participants provided references to data, along with annotations. Consequently, each pair of links and annotations was analysed as a single unit.

The data associated with each different type of manipulation was reviewed independently. For example, content identified as 'attention-grabbing' was analysed independently of content identified as 'financial gain'. This allowed for the establishment of multiple sub-categories of each of the types of manipulation identified in Chapter 3, which in turn was used in the development of the framework for understanding manipulation, which forms one of the largest contributions of this thesis.

4.4 Interpretation

This section discusses the results of the analysis of the collected data. Previously-identified types of manipulation, identified in Section 3.6, were explored and refined, with the result being that the existing classification of manipulation was made more precise.

4.4.1 Attention grabbing

Attention grabbing was the most frequently reported type of manipulation, and provided a rich variety of codes during analysis. Sub-types of attention-grabbing manipulation were identified, which were then grouped together during a second analysis pass.

The various different types of attention-grabbing manipulation are now discussed. The higher-level themes are then described and discussed.

For each type of attention-grabbing manipulation identified in the collected data, the number of items that were matched with this code are provided. Note that individually reported posts had multiple codes associated with them, which means that the sum of each type is more than the number of items reported as "attention grabbing".

Appealing and begging Posts and comments that directly prompt other users to upvote the post, visit the link or otherwise directly appealed for user attention were identified:

> *"I'm 17 and have been saving for months to be able to go to the studio with my band, it would mean the world for you to listen to the result*[4]*."*

Offensive Several posts used offensive language, or statements designed to cause offence. These included both the use of swear words in the post title, as well as linking to content and images that contained statements intended to be sexist, racist, or otherwise offensive.

One prominent example of this included a post with the title *"A friend of mine dropped* [said] *this last night after helping a woman fight off the two guys who assaulted her*[5]*."*; the post linked to an image of a man standing near a truck, with the superimposed text, *"Women shouldn't be in brawls, they belong in the kitchen*[6]*."*

[4] http://www.reddit.com/r/Music/comments/1oyyez/im_17_and_have_been_saving_for_months_to_be_able/, active as at October 24 2014

[5] http://www.reddit.com/r/AdviceAnimals/comments/1otuiq/a_friend_of_mine_dropped_this_last_night_after/, active as at October 24 2014

[6] http://imgur.com/r/all/g31nIrX, active as at October 24 2014

It is important to note that offensive language in the context of attention-grabbing manipulation is distinct from offensive language or statements in the context of attempting to disrupt the community as a whole, which was deliberately excluded from the scope of this study (as described in Section 3.3.2). Offensive language and behaviour that is intended to disrupt the wider community is not an attempt to curry favour within that community; instead, it seeks to interfere with the community as a whole.

Faking stupidity In several cases, posts and comments were reported as 'attention-gathering', along with a comment that indicated that the participant who reported on the post believed that the poster was deliberately pretending to be un-intelligent in an effort to gain attention. In one example, a post from an amateur programmer who posted a query that revealed a significant security issue in their code[7] was reported as attention-grabbing; the participant who reported the post included the comment, *"Nobody could be this stupid, right?"*

Misleading Several posts were reported as attention-grabbing that had a misleading, inaccurate or incorrect title. These included linking to news articles whose headlines did not match the news being reported; additionally, the data included posts that were flagged as "NSFW" ("not safe for work": pornographic, risque, or generally something that one may not want to be discovered looking at), despite not actually containing not-safe-for-work material. On Reddit, NSFW content is prominently badged, to avoid accidental access in public environments.

[7] http://www.reddit.com/r/PHP/comments/1l7baq/creating_a_user_from_the_web_problem/, active as at October 24 2014

Reference to self Multiple posts made references to the individual posting the content, and used first-person pronouns. These posts typically referred to items that the poster had made, or were activities that the poster was participating in.

> *"My homemade mermaid costume![8]"*

Exclamation marks Posts that contained exclamation marks were marked as attention grabbing. The majority of reported posts that contained exclamation marks contained only one exclamation mark; it was rare for reported posts to contain multiple marks.

> *"PANTHOR![9]"*

Asking for feedback Posts and comments that included a question that requested support from the reader:

> *"Bought my gf a dress and convinced her to jump into the middle of a kelp forest. What do you think?[10]"*

Attention-grabbing posts also used statements indicating that the poster believed the content of the post to be high value:

> *"This clothing store has a -20C freezing chamber where you can test their winter gear. I think it's genius."*

[8] http://www.reddit.com/r/pics/comments/1pc31n/my_homemade_mermaid_costume/, active as at October 24 2014

[9] http://www.reddit.com/r/funny/comments/1p4fgg/panthor/, active as at October 24 2014

[10] http://www.reddit.com/r/pics/comments/1p3zvo/bought_my_gf_a_dress_and_convinced_her_to_jump/, active as at October 24 2014

New thread, old topic Posting a reply to another thread as another post, rather than as a comment. Several cases of this involved the word "fixed": that is, a user would post their opinion, and another user would create a second post in reply to the first, using the same title but with different content, and appending the word "fixed" in parentheses.

Caps Several posts made use of capital letters, either for the majority or entirety of the post's title:

> *"CAN'T STOP LAUGHING: HE SAID HIS DANCING SKILLS WILL WIPE OFF ANY FROWN & I HAD NO IDEA HE PLANNED TO DANCE LIKE THIS... IN THE MUD. HILARIOUS, LAUGHING STOCK!!!*[11]*"*

Reference to Reddit Attention-grabbing posts made references to Reddit itself, often directly addressing the audience. As noted by Debevec and Romeo (1992), references to the reader using second-person pronouns (i.e. "you" and "your") encourage favourable attitudes and intentions; it is reasonable to infer that this effect extends to the use of the word "Reddit" as a similar means of collectively addressing the wider Reddit community as a whole.

For example, the following text was used as the title of a link to a picture of a puppy:

> *"Hey Reddit. Meet Akira! I think she belongs here."*

During the analysis of the data presented in this chapter, the investigators noticed several posts that, while being posted on Reddit, linked to

[11] http://www.reddit.com/r/funny/comments/1p4a8k/cant_stop_laughing_he_said_his_dancing_skills/, active as at October 24 2014

content hosted on alternative social media sites, which had post titles that targeted *both* sites.

An illustrative example of this involves a gallery of pictures of aquaria, which was submitted to the */r/pics* subreddit. Because Reddit does not host images, users who wish to share images do so via third-party services. In the case of image hosting, a popular option is Imgur (Imgur Inc. 2009, pronounced "imager"). Imgur allows users to upload and share pictures, and also serves as a social media site in its own right: users can create accounts, vote on content, and participate in an image-oriented social media community.

In this case, a link to the image gallery was submitted to Reddit, titled:

> *"Does Reddit like fish tanks?*[12]*"*

However, the title of the image gallery as hosted on Imgur differs by one important word:

> *"Does Imgur like fish tanks?*[13]*"*

The post on Reddit directly addresses the Reddit audience, and the image gallery on Imgur directly addresses the Imgur audience. In this case, the poster was customising both posts for the target site, by making specific reference to the site's name. One user who noticed this manipulation made this commented on the Reddit thread:

> *"Looks like someone is.. fishing for attention.*[14]*"*

The author disclaims all responsibility for any puns made by third parties.

[12] http://www.reddit.com/r/pics/comments/1vaoc1/does_reddit_like_fish_tanks/, active as at October 24 2014

[13] http://imgur.com/gallery/tX4Pe, active as at October 24 2014

[14] http://www.reddit.com/r/pics/comments/1vaoc1/does_reddit_like_fish_tanks/ceqivgg, active as at October 24 2014

Reference to relationships Additionally, attention-grabbing posts made specific reference to a person known to the poster, such as family or friends. This type of manipulation may be considered a variant on the "reference to self" manipulation type.

> *"I complained to my wife that she has way more karma than I do, she sent me this at work today.*[15]*"*

Long text Posts that had a longer-than-average title were reported. Because of the length of these posts and the large font size used on Reddit for post titles, they appear larger when displayed on the site, and consequently are more visually prominent than their neighbouring posts.

A particularly prominent example of this technique is this especially long post, whose entire post title, which was displayed in the site's default large, blue typeface, follows:

> *"Opration pickup Max: I'm driving my way to go pick up my son Max. From Cali to Wisconsin, and I'm about 4 hours away from Salt Lake City. I remember reading someones post that a fellow Reddit fan helped with them to crash, so let's see of this makes the front page, if not I'm sleeping in my car.*[16]*"*

Link bait Posts whose titles were phrased to be deliberately enticing and often hyperbolic, while being vague as to the precise reason for why the content was worth the readers' time.

[15] http://www.reddit.com/r/pics/comments/1p2zd4/i_complained_to_my_wife_that_she_has_way_more/, active as at October 24 2014

[16] http://www.reddit.com/r/pics/comments/1qsh2z/opration_pickup_max_im_driving_my_way_to_go_pick/, active as at October 24 2014

4.4.1.1 Spam

Posts and comments reported using the data collection tool as spam revealed an interesting trend: most items flagged as spam were not actually commercial messaging, but rather low-content contributions (*"have an upvote"*, *"wat"*, and so on.)

Some examples of these kinds of posts marked as spam include:

> *"kinda wanna give you gold for this comment."*
>
> *"Well written."*
>
> *"haha i like that analogy."*

In these cases, along with many similar examples, the comments do not contribute to the discussion, but are rather content-free statements of approval that do not contribute any opinions or information to the conversation beyond that of the poster's approval. While these comments technically contravene Reddit's site-wide community guidelines - specifically, the section that requests users not post content-free comments (Reddit Inc. 2014a) - the element of most interest to this research is the fact that several items classified as "spam" by participants are not advertising any product or service.

This suggests that users of Reddit, and potentially other social media communities as well, have a different definition of spam: one that emphasises the lack of relevant content, rather than any commercial intent.

As a result of this finding, the follow-up interviews with participants were designed to further explore this, and gather additional information on user perspectives of spam. This is explored in more detail in Section 5.3.4.

This section has discussed multiple specific subtypes of 'attention-grabbing' manipulation. It is important to note that this is not intended to be a complete list of all possible types of attention-grabbing manipulation; rather, this thesis presents the types of attention-grabbing manipulation identified during this study. Attention-grabbing manipulation was the most prominent type of manipulation identified by participants in this study.

4.4.2 Financial gain

Financial gain, in the context of study, means posting links to content in an attempt to market a product or service.

Posts and comments that were flagged by participants as being examples of financial gain were found to fall into two categories:

- Posts that promoted a product or service created by the individual poster themselves, and
- Posts that incorporated branding elements from a corporation.

Posts that promoted a product or service tended to intersect with the *reference to self* and *asking for feedback* types of manipulation, but included either explicit or indirect requests for users to make a purchase.

> *"My husband has spent 4 years of his spare time making this album. It's his first, and I think it's pretty good. I'd love if you listened!* [17] *"*

Posts that incorporated branding elements from corporations tended to not include any reference to a product or service, but rather otherwise unrelated posts that happened to reference a commercial brand.

[17] http://www.reddit.com/r/Music/comments/1otbwv/my_husband_has_spent_4_years_of_his_spare_time/, active as at October 24 2014

Examples included users posting links that mentioned commercial brands, or employees of large companies, in a positive light.

> *"KFC gets it wow* [18]*"*

> *"This is why I love the Hampton Inn* [19]*"*

> *"Old school Coca-Cola ad* [20]*"*

> *"Wendy's employee removes umbrella from table outside to walk elderly gentleman to his car in the rain. Faith in humanity partially restored.* [21]*"*

Participants also labelled as "financial gain" a small number of posts that more accurately suited the definition of spam used in this research: that is, direct links to sales of products that were neither being promoted directly by the individual who created them nor were indirect marketing. These posts were re-classified as spam and analysed along with the other data coded by users as such.

4.4.3 Personality voting

Personality voting, in the context of this research, refers to a manipulation technique that places the focus of the content on the identity or circumstances of the individual who posted the content. This is distinct from

[18] http://www.reddit.com/r/SuperShibe/comments/1ic1tx/kfc_gets_it_wow/, active as at October 24 2014

[19] http://www.reddit.com/r/funny/comments/1p6zrx/this_is_why_i_love_the_hampton_inn/, active as at October 24 2014

[20] http://www.reddit.com/r/pics/comments/1p48co/old_school_cocacola_ad/, active as at October 24 2014

[21] http://www.reddit.com/r/pics/comments/1osayd/wendys_employee_removes_umbrella_from_table/, active as at October 24 2014

"reference to self" attention-grabbing manipulation (Section 4.4.1), in that the essence of "reference to self" is "you should care about this subject because I am involved with it to some degree", while that of "personality voting" is "you should care about me as the subject in my own right".

Posts categorized as "personality" voting fell into three main categories:

"I'm special" Posts that made direct reference to either the identity of the poster, or to an attribute of the poster. Examples included a post by former Governor of California and actor Arnold Schwarzenegger titled "I'm back", and a post that made reference to the poster's "real-life cakeday" (that is, the poster's birthday).

Sympathy vote A subcategory of "I'm special": Posts categorised as "sympathy vote" were those that drew attention to an attribute of the poster (such as a disability or illness), which the poster asked for sympathy or support for.

Cakeday A subcategory of "I'm special": Posts that made reference to the user's "cake-day", which is the anniversary of the date the poster created their Reddit account.

4.4.4 Requesting upvotes

Posts that requested upvotes from other users were found to fall into two categories: those that *implicitly* requested upvotes, and those that *explicitly* requested upvotes.

Implicit requests for votes either used questions (*"How did we do?"*, or *"Why hasn't this gotten more attention?"*), or were comments that stated that the post should be upvoted (*"You need more upvotes, this is heartwarming."*;

"This deserves way more upvotes.")

4.4.5 Rewarding upvotes

Rewarding upvotes took the form of offering a financial reward, generally for a charity; one prominent example of a single user being rewarded for votes involved this exchange between a Reddit user and team of Microsoft product developers working on a new model of tablet computer:

> *"***[User:]** *Can I have one? (Doesn't hurt to ask, right?)"*
>
> *"***[Developer:]** *Get this voted to the top, and you got it."*

The user's comment received enough votes to place it at the top of the page. Approximately 25 minutes after their offer, the developer commented:

> *"***[Developer:]** *To quote one of my favorite movies, Tin Cup: "Winner, winner chicken dinner!" DM* [direct message] *your details and I'll send you your Surface Pro 3.*[22]*"*

4.4.6 Organising mass votes

The majority of data reported as 'organising mass votes' took the form of either talking about the user's content being downvoted, or discussing upvoting content in a different area of the site.

[22] http://www.reddit.com/r/IAmA/comments/26m9cu/we_are_panos_panay_and_the_surface_team_at/, active as at October 24 2014

4.5 Discussion

This section discusses the implications of the findings presented in Section 4.4, and presents the contributions of this chapter: a refinement of the framework of manipulation that was presented in Chapter 3, and an evaluation of the web-based data collection tool used in this study.

4.5.1 Web-based data collection

The web-based data collection tool used in this study was extremely successful in gathering useful data, while operating under a number of constraints.

The design constraints for the tool were:

- *Public access to the tool*: In order to encourage participation from as many people as possible, the tool was made available to the public. Participation did not require any communication or interaction with the researcher; additionally, participants were free to recruit other individuals to the study.

- *Minimal impact on normal browsing experience*: The data collection tool was required to minimise the interruption of the normal experience of browsing Reddit. Accordingly, the only change that the tool made to user's experience of Reddit was the addition of a single link to the pre-existing collection of management links present under each post and comment.

- *Real-time data collection*: In order to ensure that all reported data was received for analysis, and to avoid problems associated with storing information for later colletion (such as data loss, or participant

delinquence), the data collection tool transmitted all information immediately as it was report.

The data collected through this tool proved straightforward to analyse. Because all data was received in a standard format, the data was able to be easily collated for analysis. The analysis in this study made use of the fact that the data was partitioned into the different types of manipulation identified in Chapter 3; this was a direct consequence of the fact that the user interface for the tool required users to select the overall type of manipulation that they were reporting.

The fact that the web browser extension component of the data collection tool was able to be remotely updated by the researcher allowed for rapid updates to issues as they were identified. This was further improved by the fact that information was received by the server component immediately, which allowed for extremely rapid reaction times for any critical errors. Although no critical errors were experienced, the author is confident that if any had occurred, the ability to quickly provide an updated version of the software would have resulted in minimal loss of participant-reported data.

Finally, the author was able to get a feel for the data early in the study, as data came in. While data analysis did not commence until data collection had entirely ceased, the fact that data arrived continuously meant that the author was able to spend more time familiarising himself with the data.

4.5.2 Extending the framework

An immediately important item to note is that no posts or comments were submitted that did not fit within the scope of the types of manipulation

identified by administrators and moderators. This suggests that the initial model of manipulation did not contain significant omissions in its broad classifications of manipulation: participants were free to provide additional commentary, and none complained of any lack of appropriate cateogry for the content that was being reported.

Additionally, more precise terms for the different types of manipulation were discovered. The 'attention grabbing' type of manipulation was the most diverse, with an additional thirteen sub-types of attention-grabbing manipulation.

Each additional sub-type of manipulation, therefore, may be used to add to the nascent framework of manipulation, presented in earlier chapters. These additions proved instrumental in gathering information on the impacts of manipulation, the report concerning which is presented in Chapter 5.

The extension of the framework is shown in Figure 4.3.

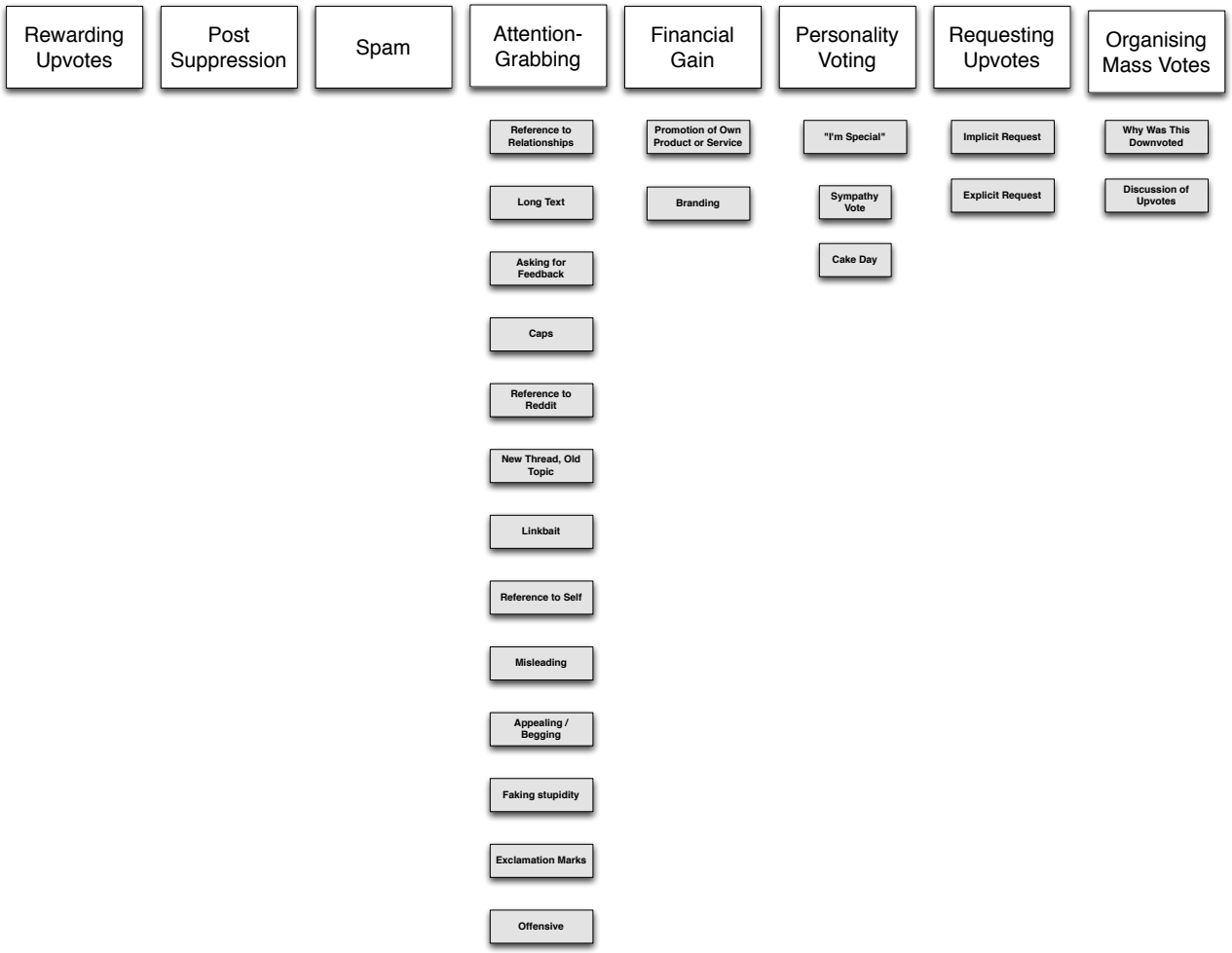


Figure 4.3: The extended framework of manipulation, presenting the sub-types of manipulation identified in Phase 2.

In Chapter 3, a definition of manipulation was established. To reiterate this definition: *manipulation on social media sites is an attempt to change the global algorithmic relevance of content by changing its individual topical and affective relevance*.

The study described in this chapter has explored the different types of manipulation initially identified in Chapter 3, and has identified sub-categories of several of these types. Having extended the categories of manipulation, these categories are now re-examined in the context of the above definition. Specifically, each category and sub-category of manipulation identified in this chapter can be examined and the form of relevance that it is attempting to influence can be named. This is presented in Figure 4.4.

Manipulation Type	Type of Relevance	Justification
Rewarding upvotes	**Affective / Algorithmic**	Either direct reward for up votes, or satisfaction for helping a cause
Spam	**Topical**	Pretends to match topics the user is seeking
Attention-grabbing		
Reference to Relationships	**Affective**	Satisfaction from providing exposure to a follow community member's friend / relation
Long Text	**Topical**	Seeks to provide as much information as possible to establish topicality
Asking for Feedback	**Affective**	Seeks to satisfy user desire to provide opinion
Caps	**Affective**	Seeks to provoke an emotional reaction through interpreting the post as "shouting"
Reference to Reddit	**Affective**	Seeks to satisfy user feelings of community engagement
New Thread, Old Topic	**Topical**	Seeks to increase likelihood of the content matching the topic the reader is seeking
Linkbait	**Affective**	Seeks to provoke feelings of satisfaction through outrage, sympathy, curiosity
Reference to Self	**Affective**	Seeks to provoke feelings of community support
Misleading	**Cognitive**	Pretends to be new information
Appealing / Begging	**Affective**	Seeks to provoke feelings of community support
Faking stupidity	**Affective**	Seeks to provoke feelings of satisfaction through amusement at the poster
Exclamation Marks	**Affective**	Seeks to provoke an emotional reaction through emotive punctuation
Offensive	**Affective**	Seeks to provoke a reaction, often humour
Personality Voting		
I'm Special	**Topical**	Hopes to transfer interest in poster to interest in topic
Sympathy Vote	**Affective**	Satisfaction from "helping someone out" via vote
Cake Day	**Affective**	Satisfaction from acknowledging someone's commitment to community
Requesting Upvotes		
Implicit Request	**Affective / Algorithmic**	Seeks to provoke a feeling of sympathy for the poster's lack of recognition
Explicit Request	**Algorithmic**	Seeks to directly increase the vote count of the content
Organising Mass Votes		
Why Was This Downvoted	**Affective**	Seeks to provoke a feeling of sympathy for the poster's reception
Discussion of Upvotes	**Algorithmic**	Encouragement of direct changes to the votes for content

Figure 4.4: The extended framework, with types of manipulation identified.

4.6 Summary

This chapter has reported on the web-based data collection study that comprises the second phase of the research. In developing a web-based tool for gathering user perspectives on manipulation, and the relative prevalence of different types of manipulation, a significantly more thorough exploration of manipulation was made possible.

This chapter has achieved the objectives stated in Section 4.2.1 in the following ways:

- Attention-grabbing manipulation has been identified as the most prevalent form of manipulation, relative to the other types identified in Chapter 3, and was found to have the widest range of different interpretations by end-users. It is therefore justifiable to claim that attention-grabbing manipulation is likely to be the type of manipulation that has the greatest impact on users.
- A second-round analysis of the framework in Chapter 3 has demonstrated that the framework holds up when given to a wider audience for evaluation. No data was submitted that was found to not fit within the framework proposed.
- A web-based data collection tool was developed and evaluated; the software worked well for its intended purpose, and the researchers intend on continuing to use this approach for in-situ data collection purposes.

Having explored the various forms in which manipulation was encountered by users of the site, it became possible to now consider the impact of this manipulation on the user's use of the site. Chapter 5, following, discusses this at length.

5

Phase 3: User interviews

This chapter presents a discussion of Phase 3 of the overall research project: a series of interviews with users of Reddit, which aimed to gather information about what impact different types of manipulation have on users of the site.

In these interviews, users of Reddit who had previously participated in the earlier study were interviewed about their views on manipulation, on how various types of manipulation affected their use of Reddit. Participants were also asked questions regarding their use of Reddit in general, including what kinds of content caused them to continue reading the site, and what kinds caused them to stop reading.

5.1 Introduction

The third phase of the research presented in this thesis sought to create an understanding of how the various types of manipulation, which were identified in Chapter 4, affected the users of social media sites.

In previous chapters, this research had confirmed several underlying factors:

- Chapter 3 had determined that manipulation is seen by moderators as an issue. Additionally, Chapter 3 identified the different kinds of manipulation seen.
- Chapter 4 confirmed that users of the site also see manipulation, and provided a finer-grained analysis of these different kinds.

The key element missing from these two previous studies, which is now addressed in this chapter, is a lack of understanding of how different kinds of manipulation affect the users of the site. While previous chapters identified that manipulation was a problem that took multiple forms, no attempt had yet been made to determine the various ways that manipulation changed the behaviour of users of the site.

It is important to note that the study presented in this chapter was not intended to attempt to quantify the impact of manipulation on users; rather, by gathering viewpoints from site users on the various kinds of impact that manipulation has on them, the research enabled the construction a framework for addressing those impacts.

5.1.1 Chapter Structure

The structure used by this chapter is as follows:

- Section 5.2 discusses the approach taken for the study presented in this chapter, presents the overall methodology and objectives, and discusses ethical considerations.
- Section 5.3 presents the findings of this chapter, and discusses their implications for the research presented in this thesis.
- Section 5.4 discusses the findings of the chapter in the context of the research as a whole, and presents the conclusions.

- Section 5.5 summarises the work presented in this chapter.

5.2 Approach

In order to gather information about how users are affected by manipulation, a more in-depth approach to gathering information was required: while the web-based data collection tool used in Chapter 4 served well in gathering a broad collection of information, generating deeper understanding of the effects that manipulation had on users required a deeper approach.

The approach taken in this third phase mirrored that of the first phase, discussed in Section 3.2. Semi-structured interviews were carried out and the results analysed, following the same methodology of both data collection and analysis as used in Chapter 3.

The structure of this section is the following:

- Section 5.2.1 discusses the objectives of this study.
- Section 5.2.2 discusses the ethical precautions taken during the study.
- Section 5.2.3 discusses the contributions made by this chapter.
- Section 5.2.4 presents the methodological approach used to collect the data.
- Section 5.2.5 discusses the recruitment process for the study, and describes the participants.

5.2.1 Objectives

The overall objective of this study was to collect specific, detailed information about what different impacts manipulation has on the users of a

social media site (namely Reddit), in support of the overall goals of the research defined in Section 1.2. Alongside this overall goal, the study was also intended to gather user perspectives on the web-based data collection tool used in Chapter 4.

In order to gather specific, useful information about the impact of manipulation on users, specific objectives were indentified:

- *to determine the specific actions that users took when they saw manipulation*, and
- *to determine the severity of the impact that manipulation had.*

The author began this study with several presumed actions that it was felt users were likely to take, such as down-voting manipulating content; however, this was likely to be incomplete and lacked grounding in the data.

It was also felt by the author that different kinds of manipulation would have varying degrees of impact on users. While no attempt to quantify this impact was made, an understanding of how harmful certain manipulation techniques were was a desired outcome.

By validating the data collected in earlier chapters and by creating a deeper understanding of user behaviour on the site covered in this case study, the finalisation and validation of the framework of ranking manipulation was made possible.

5.2.2 Ethics

As with the interviews conducted in phase 1 of the research, the following precautions were taken when conducting the interviews presented in this chapter:

- Participants were not asked for any personally identifying information. When analysis began, all names, usernames and other pseudonyms were removed.
- Participants were free, at any time, to withdraw from the study, prior to the data collection being completed.
- Participants were informed that no judgements were being made regarding their behaviour on the site, or their interests or preferred topic.

The study described in this section was approved as a Minimal Risk Study by the Tasmanian Human Research Ethics Committee. The reference number for this study was H0013714.

5.2.3 Contributions

This chapter makes the following contributions to the overall research:

- The types of impacts that different kinds of manipulation have on site users is identified.
- The severity of different impacts, gleaned from conversations with users, is identified and discussed.
- A more complete evaluation of manipulation, and the various forms it takes.

5.2.4 Design

Data for the study presented in this chapter was gathered through a series of semi-structured interviews. Semi-structured interviews, as has already

been noted in Section 3.3.1.1, have proven an excellent tool for collecting 'deep' information about a topic. The reader is directed to Section 3.3.1 for a more detailed discussion of semi-structured interviews, including their strengths and weaknesses.

A fine-grained, systematic approach to the analysis of the data collected in this study was required; as Kjeldskov and Graham (2003) notes, the rich, natural character of data collected through semi-structured interviews requires careful analysis in order to overcome the often contradictory nature of the information gathered. Accordingly, a grounded approach (as per Corbin and Strauss 1990) was used to analyse the collected information, and derive the findings presented in Section 5.3.

5.2.5 Data collection and participation

In order to gather the information required to satisfy the objectives laid out in Section 5.2.1, careful selection of the focus questions used in the interview was required. The focus questions for the interview were designed around three themes:

1. *Use of Reddit*: Participants were asked about their general use of Reddit, including the topics they prefer, and how frequently they post, comment and upvote (and indeed whether they post, comment or upvote at all).

2. *Reactions to manipulation*: Participants where then asked to describe how they reacted to manipulation. The classification of types of manipulation (both in its original form from Chapter 3 and the elaborated framework derived in Chapter 4) proved invaluable in the design of these questions, by providing a number of specific prompts for questions. For example, by allowing the investigators to specif-

ically ask about the participant's reactions to the use of all-caps in post titles, more detailed information about the impact of manipulation could be uncovered.

3. *Impact of data collection tool*: In recognition of the fact that an 'observer effect' is possible when running a diary study, participants were asked to reflect upon what effects the data collection tool may have had on their use of Reddit.

Examples of these focus questions include:

- How long have you used Reddit for?
- What do you generally use Reddit for?
- What would you say makes a post worthy of an upvote?
- When people reply to a post with another post, as opposed to replying with a comment, does that affect the way you browse the site?
- When you see a post made by someone that you recognize, for example celebrities and things, are you more likely to look at those kinds of things?
- What are your thoughts on people posting and asking for sympathy?

Participants were recruited through an open invitation included in the announcement of the Phase 2 study (discussed in Chapter 4). Interested users were asked to reply to the announcement post; they were then contacted by the author, using Reddit's internal messaging system. Five users were recruited in this manner. Three users were able to be interviewed face-to-face, while the remaining two were interviewed over instant messaging.

As with the interviews conducted in Chapter 3, the number of participants was not intended to be statistically significant; the purpose of these interviews was to gather information about different types of impact, rather than the frequency with which these impacts were noticed. Reddit's diverse community means that it is possibly that the heterogeneity of the population is entirely captured by the sample size, but the author felt that the participants provided sufficient perspective and valuable insight for the goals of this phase.

The face-to-face users were interviewed in the office of the author, and each interview took around 30 minutes to complete. Interviews were transcribed, and all participants were male. All participants had a similar level of Reddit experience, and had been using Reddit actively for more than a year in all cases but one, who had been a Reddit user for around 8 months.

5.3 Analysis

The analysis of the data collected in in the course was analysed using similar techniques to those used in Chapter 3 and Chapter 4. A grounded-theory-inspired approach was used to first reduce the volume of data to one that could be effectively analysed; following this process, open coding was used to determine low-level themes. These were then grouped into higher-level themes via thematic coding. More information on these processes can be found in Section 3.5.

This section details the findings of the study, and discusses how they integrate with the rest of the thesis.

In the course of analysing the interviews conducted with participants, the researchers noticed an interesting spread in the severity of the types of impacts that manipulation has on users. Some forms were universally

reviled; others were treated with ambivalence, and some forms were actively enjoyed and found to be useful by participants.

The process of thematic coding discussed in Section 5.2.4 resulting in the identification of three overarching themes:

- *Direct* impacts;
- *Indirect* impacts; and
- *Neutral/positive* impacts.

The remainder of this section discusses each of these categories in detail.

5.3.1 Direct impacts

Direct impacts were the direct actions that users took as a result of perceiving manipulation. Direct impacts have an immediate effect, either on the user, or on the content (due to actions taken by the user with the content.) Consequently, these direct impact are considered to have a high degree of severity.

Skipped over / ignore Participants reported that, in serveral cases, seeing manipulation caused them to begin ignoring content that they would otherwise have looked at. This was the most prevalent of all kinds of impact: all participants reported having ignored content on the basis of a post title that they viewed as attempting to be manipulative.

A good example of this is one participant who, when asked about their response to attention-grabbing post titles, replied:

> *"I completely ignore them. You get enough crap on every other social media or website."*

Down vote Down-voting content is the most basic action that users on Reddit can perform. As was discussed in the overall discussion on how Reddit works in Section 2.4.4, Reddit users may indicate their approval for content by up-voting, and their disapproval by down-voting.

Users reported downvoting in two context: first, as a matter of course, and secondly, with a specific outcome in mind. Users who described downvoting as a natural part of their use of Reddit did so as part of their usual purusal of the site. One participant, in describing their reaction to re-posted or duplicate content, described a typical example of this pattern:

> *"I generally just downvote unless it's a particularly witty twist on [a repost]. Then I leave it be."*

Interestingly, many participants reported that they tended not to downvote at all. Two reasons for this were identified: either the participant not being in the habit of downvoting, or the fact that the participants were often not signed in, which is required in order to vote.

Users who were not in the habit of downvoting tended to describe themselves as simply someone who doesn't downvote content:

> *"I don't remember the last time I downvoted."*

Another participant reported only one memory of down-voting:

> *"I think I've only down voted once, for a spammer. Yeah, I don't really down vote much."*

Users also reported wanting to downvote, but were not signed in to the site at the time, and did not wish to sign in. One participant noted that he fell out of the habit of down-voting content as a result of an extended period of time of not being able to:

> *"I used to up vote and down vote things, and I just sort of stopped, because I started switching between computers a lot, and I wasn't logged in for a few times, and there were only a few subreddits I was reading at the time, so I just went to them manually.*
>
> *I never logged in, and now that I actually have gone back and logged into the computers that I use, I just don't, for some reason. I think it's because, when I went to up vote a few things, Reddit says "you have to be logged in to do that", and I just can't be bothered. It was kind of odd. I started off doing it, and I just stopped."*

Unsubscribe / Leave reddit In several cases, participants reported manipulative content causing them to leave Reddit, or to leave a sub-community of Reddit.

A participant who was discussing marketing posts (see Section 3.6.7), described this action:

> *"[I] just skipped over it, ignored them, left the reddit for the day and came back the next day."*

Other instances of departing a community had longer-lasting effects:

> *"Everyone would post pictures of stuff that their girlfriends and whatever had given them. Just the same crap. Everyone got the same Hylian shield* [an item from *Zelda* video games] *for christmas. I really don't care about that, and so I don't look at the Zelda subreddit. I stopped looking at the Zelda subreddit for a month... it actually changed the way I browse that subreddit, in that I don't anymore."*

Hide Reddit provides a mechanism that allows users to hide a post, which prevents them from seeing it appear when later browsing the site.

A participant mentioned that, when he saw posts comprised of all-capital letters (see Section 4.4.1), he would almost always hide them:

> *"If it's a fairly low-traffic subreddit, like some of the board game ones are, something that's going to annoy me for a few weeks like the all-caps thing, I'm more likely to hide it."*

Tagging While not a built-in feature in the Reddit sites, several third-party add-ons for the site - notably the Reddit Enhancement Suite by (Sobel 2014) - allow users to associate posts or users with a custom label, for future reference.

> *"I tend to chuckle, smile, and usually tag them as being an idiot and disregard them later on."*

5.3.2 Indirect impacts

Indirect impacts were those that did not involve a direct action being taken by the user upon identifying manipulation, but had indirect effects on their opinion of the content or of the community in which it was posted.

Disappointment Participants reported a feeling of disappointment when they found highly-ranked content that did not meet their expectations. The underlying reasons for this varied from instance to instance, but all participants reported a common feeling of being misled by the poster.

One user reported that the "acting stupid" manipulation type led to an overall reduction in their participation in a community:

> *"I'm always a bit disappointed when I see it in one of the subs that I think more highly of, like the discussion subs.*

> *I recognise there are exceptions to every standard. However, if a sub is consistently bad, then I don't come back. I think /r/PoliticalDiscussion is a good example of this. It's supposedly a political discussion sub, but I find that it tends to only be a small step above something like /r/Politics, so I avoid it."*

Suspicion Participants also reported feeling suspicious of posts, and spending time thinking about whether or not the content should be downvoted (or having any of the other actions discussed in the *Direct Actions* section above).

> *"I don't downvote because, well, I can rationally recognise that they're pandering, but I also don't want to be a jerk. I just leave them."*

Annoyance Participants frequently made reference to the fact that manipulative content made them annoyed. The intensity of this annoyance different between participants, but every single participant made at least one reference to this kind of impact.

One participant, while discussing "cake-day" attention-grabbing posts (see Section 3.6.5), described the feeling:

> *"I do get annoyed when there's content like, "it's my cakeday, here's a picture of something". I don't click on that, because I don't see the point. Again, because it's certain date on which you signed up for Reddit, doesn't make you better at Reddit. It doesn't make you a more interesting person. It doesn't make your content better, so why should I click on it?"*

Obnoxious/blatant Many participants felt that manipulative content affected them more when the technique being used was more obvious.

> *"It's almost like a weird cry for attention. Whereas, if it was more along the lines of, "I made a spaceship with 93 engines in a circular pattern so it explodes when it takes off*[1]*", I'd be more intrigued to click on that than I would be for a post titled "hey guys, look at this exploding rocket I made." I think that might just entirely be due to the phrasing."*

Not interesting Participants found themselves less interested in manipulative posts, and in subreddits that frequently contained manipulative posts.

One participant mentioned that he tended to cease his browsing of reddit when *the content becomes polluted with things that I don't normally look at.*

A user reported leaving Reddit for a time:

> *"A day or two, until I remembered to look again. After I'd gotten over the annoyance of having to look at people going, "this is my baby, look how cute he is; it's my cake day, vote me up.""*

5.3.3 Neutral/positive impacts

Several types of manipulation identified in the previous phase were identified by users as having either a neutral impact on them, or having a positive effect. This was surprising to the author, and its implications are discussed in more detail in Section 5.4.

[1]This participant was referring to the Kerbal Space Program subreddit: http://reddit.com/r/kerbalspaceprogram

Not a problem Several users simply reported that certain kinds of manipulation were not a problem to them, and had no impact on their browsing of the site.

One participant, while discussing long post titles, noted that he did not mind long titles, as long as they were not unneccessarily long:

> *"If the post title is humorous or if it's serious, having it be long isn't really a problem."*

This statement was echoed by another participant:

> *"Usually the more stuff in the post title means I don't have to spend quite so long investigating what the link is about, and it doesn't seem to be some sort of shitty buzzfeed titled link."*

Indeed, some kinds of 'manipulation' previously identified were even described as necessary to the operation of the site. For example, one user commented on duplicated ("re-posted") content:

> *"The userbase is constantly changing. There's no harm in them, especially when people haven't seen it before... Reposts are necessary. Not everyone gets to see things the first time around.*
>
> *For lots of people, they're new content, otherwise they wouldn't have been upvoted in the first place. But in terms of their effect on the site as a whole? It's better to have entirely new content, but then, there's so much of that that reposts don't really have a particularly big effect."*

Click it / upvote Some participants noted that certain kinds of manipulation made them *more likely* to view the content, or to upvote it.

This sometimes occurred even when the participant disliked the type of manipulation being used. One participant, discussing posts to images that included requests for votes, noted:

> *"I look at most pictures. Pictures are easy to look at."*

5.3.4 Definitions of spam

During the course of the interviews, one of the questions asked of the participants by the researcher regarded the participants' personal definition of spam. The inclusion of this question, as was foreshadowed in Section 4.4.1.1, was prompted by the fact that many of the items flagged as spam by participants in the study described in Chapter 4 appeared to not actually be advertisements of any kind, but rather were terse, low-content comments.

When participants were asked to define spam, the predominant theme was that of spam meaning 'low quality content':

> *"I'd like to go so far as to say something that adds absolutely nothing to a conversation. But that's probably like half of Reddit. For me it's hard to define spam, because lots of things have worth, but just because I don't see it as having worth, doesn't mean it's spam."*

One participant noted that their definition of spam was a deliberate extension of the popular definition:

> "[Spam is] *unwanted content, I would say... if a subreddit is about turtles, posting a thread about cheese would be spam, because it has nothing to do with the subreddit, it's not applicable in the slightest. It doesn't necessarily have to be about finance, as in financial gain*

for the spammer, but its content has nothing to do with the thing at hand. So, as such, it's spam."

This indicates a potential drift in definition of spam: while the spam is characterised by Brunton (2013) as having a shifting definition, it is generally accepted as the bulk posting of unsolicited messages (Spamhaus Project 2014). This finding - that spam's definition may be extending to include individual, low-value posts - does not lead to further information regarding the impact of manipulation, but it serves as an opportunity for further research into user definitions of spam.

5.3.5 Evaluation of the web browser extension

Participants were asked to reflect upon how the web browser extension affected their use of Reddit. Participants reported generally positive feelings about their use of the software; the following comment from an interview participant indicates a representative viewpoint:

"I'd often be like, "oh, I wonder if I should report that? Should I bother reporting that?" I wouldn't say it changed it substantially, but it did change it a little bit. You'd spend a moment or two musing over whether content was worth reporting or not.

It caused me to think more about the content that I probably wouldn't normally think about. But I go to a lot of niche subreddits."

Another participant indicated similar sentiments with regard to the software causing them to spend more time thinking about content on the website, and about content manipulation on the website:

"Yes, I was more on the lookout for things that could be considered manipulation. For a brief period, if I was bored, it gave me something to do on Reddit by searching, using the search feature to search

for words commonly associated with some of those things, so I could specifically flag them. Or just have a look at them. And I still somewhat do that, even though I don't have the extension installed at the moment.

(Interviewer: So, you now actually...)

Find things to get angry at, yes."

One participant felt that their overall experience of Reddit was enhanced as a result of participating:

"I found myself paying a lot more attention to posts and their techniques than I usually do. It was really nice, actually."

It is worth re-iterating that the web browser extension, when installed, was a relatively visible piece of software: all pages on Reddit viewed by the user were modified to some degree, and while the software was designed to not deliberately interfere with normal use, the modifications made by the software to the user's experience of Reddit were pervasive.

Nonetheless, no participants interviewed provided any negative feedback on how the software changed their experience of Reddit; indeed, as has been noted in the comments above, several found it provided a positive experience. The author is therefore comfortable in concluding that the use of the web-based data collection tool was a success.

5.4 Discussion

This section discusses and interprets the findings of the study presented in this chapter. It is worth reiterating at this point that the goal of the study presented in this chapter does not attempt to draw connections between the types of manipulation identified in Chapter 3 and Chapter 4. Rather,

the study aims to provide a broader understanding of what kinds of impacts exist.

The analysis of the data collected in this study revealed three main categories of impact types: direct impacts, which have an immediate or near-immediate effect on the user's behaviour on the social media site, indirect impacts, which have a cumulative effect on the user's use of the site, and positive impacts, which may be considered to be a sub-type of indirect impacts that have a beneficial impact on the user's use of the site.

As was noted in Section 5.3.3, participants identified several types of impacts as having a neutral or positive impact on their use of the site. This was surprising to the author; both of the previous studies indicated that content ranking manipulation was perceived to have negative consequences for end-users, by both administrators and moderators in the first study, and by the participants in the second.

This has interesting implications for the topic of content ranking manipulation as a whole, as it changes the focus of the disussion from one centered on ameliorating negative impacts to one in which the impact of manipulation upon the user must be identified as positive, negative or neutral.

5.5 Summary

This chapter has reported on the third and final phase of data collection used in this research. Broadly, this study has achieved the objectives set out by Section 5.2.1 in the following ways:

- The types of impact that manipulation can have on users have been identified, as well as their severity.

- The definition of spam held by participants has been noted to extend to *any* low-value content, and is not limited to unsolicited bulk messaging.
- The choice of methodology used in Chapter 4 has been validated by demonstrating that participants found the web browser extension to be unintrusive, and in fact improved the experience of browsing Reddit in some cases.

The following chapter, Chapter 6, integrates the findings and contributions made by the three phases of data collection, and presents the overall findings of the research presented in this thesis.

6

Discussion

This chapter integrates the findings of each of the three previous phases of the study, and discusses their implications. A model of manipulation and its impact upon social media communities is presented and discussed.

6.1 Introduction

The overarching premise of this research was that *there is insufficient research into the impact of ranking manipulation on social media sites*. In order to address this absence, the research presented in this thesis has conducted several studies that seek to create an understanding of what manipulation *is*, in the context of social media sites, as well as an understanding of how manipulation affects the behaviour of the users of those sites.

The findings of these studies have provided sufficient material to allow the construction of a framework for discussing manipulation and its effects. In this chapter, the findings from previous chapters are integrated, and the framework is constructed.

Following this, the chapter then discusses the applications of the framework in a variety of cases: from the perspective of social media site admin-

istrators, from the perspective of users who post content to social media sites, and from the perspective of developers who seek to improve the quality of the user experience of social media sites.

To review: the research presented thus far has achieved the following:

- A definition of manipulation has been created, grounded in both the perspective of administrators and users (first presented in Chapter 3 and elaborated in Chapter 4.)
- Different types of manipulation have been explored, and a classification of these types has been created (in Chapter 4.)
- The various types of impacts that manipulation have on users has been explored, and the research has found that these range from actively driving people away from the site to enhancing the user experience and clarifying it (in Chapter 5.)

6.1.1 Chapter Structure

The structure of this chapter is as follows:

- Section 6.2 discusses the research presented in Chapters 3 to 5, and the relationship between manipulation and the community in which this manipulation is situated.
- Section 6.3 discusses the relationship between manipulation and relevance, and presents possible means of addressing manipulation in the context of systemic relevance.
- Section 6.4 concludes the chapter, and reviews the work undertaken.

6.2 Understanding manipulation

This section suggests a framework for the discussion of manipulation in social media sites, taking into account the context of the community in which manipulative content is posted.

In Chapter 3, a classification of types of manipulation was identified based on the perspectives of administrators and moderators, and was explored and extended with the perspectives of users of the community in Chapter 4. These two combined studies create a broad picture of how manipulation may manifest.

As discussed in Section 5.3, not all of the manipulation types identified in Chapter 3 are viewed by users as having a negative impact, or even are manipulation at all. The definition of 'manipulation', therefore, varies between communities.

The remainder of this section has the following structure:

- Section 6.2.1 discusses various types of communities, with regards to the source of their content.
- Section 6.2.2 discusses the impact of manipulation in these communities.

6.2.1 Manipulation in communities

In order to effectively discuss what manipulation means within communities, it is useful to identify key differences between communities. There are a wide variety of different types of online community, ranging from discussion forums to image boards, through to online question-and-answer forums (discussed in Section 2.3.4). These different types of community each present their own variants on user interaction, social networking

structures, and the manner of content that is favoured by the users. In addition to these forms of social media site, there exist a multiplicity of other forms not listed here; a detailed classification of types of social media sites is beyond the scope of this research, though it is reasonable to state that variety abounds.

It is not the intended purpose of this thesis to identify and categorise types of social media site; rather, this thesis deals in social media sites as a whole, using Reddit as a case study. As has been previously discussed, Reddit serves as a particularly good case study for the discussion of social media sites, due to the fact that it serves as a platform for users to create their *own* communities.

Consequently, Reddit is a social media site comprised of a large variety of different kinds of communities, each with its own unique characteristics. While Reddit is not a question-and-answer site on its own, it contains question-and-answer communities (such as /r/AskReddit); while Reddit us not an image board, it contains image boards (such as /r/pics). This means that Reddit may be used to discuss differences between different types of communities, while still ensuring that the communities under discussion have a great deal in common with each other, due to the fact that they operate within the framework and behaviours of Reddit as a whole.

6.2.1.1 Types of communities

We may begin our discussion of the manipulation framework by dividing social media communities into two broad categories:

1. communities in which content is chiefly sourced from members of the community themselves; and

2. communities in which content is chiefly sourced from external sources.

This classification is reasonably straightforward, and we shall briefly consider it, before proceeding onwards with a discussion on to how it relates to manipulation. For brevity, the remainder of this chapter refers to the former type of community as *"created-content communities"*, and to the latter as *"discovered-content communities"*.

6.2.1.2 Created-content communities

Some examples of *created-content communities*, in which content is chiefly sourced from members of the community itself, include:

- Discussion forums, in which users pose questions to the community for discussion. Examples include:
  - */r/AskReddit*: Users pose general questions to the community at large, in the hopes of creating a discussion. Examples of typical questions posted to */r/AskReddit* include *"What makes you really angry?"*, and *"What is your best purchase of 2013?"*.
  - */r/IAmA*: Users with interesting professions or other characteristics participate in an open interview with the community.
  - */r/ExplainLikeImFive*: Users post requests for simple explanations of topics, such as *"Explain like I'm Five: What is the global financial crisis?"*
- Discussion communities for games that allow players to create their own content. Examples include:
  - */r/KerbalSpaceProgram*: Kerbal Space Program (Squad Inc. 2014) is game in which players construct space vehicles, and send them to orbit; members of this community post pictures of their constructions, and discuss design techniques.

  - */r/TheSims*: The Sims (Electronic Arts Inc. 2014) is a series of games in which players build a house, and play with a simulated family; users of this community post pictures of their houses and family, often detailing their simulated exploits and adventures.

- Creative arts communities, such as:

  - */r/drawing*: A community in which members share illustrations with one another, offering feedback and advice
  - */r/woodworking*: A community of woodworkers, in which members share photographs of their projects, as well as links to external resources.

6.2.1.3 Discovered-content communities

Discovered-content communities are those whose content is primarily sourced from external locations. Examples of these to be found on Reddit include:

- */r/news*: A news community, in which members share and discuss news stories published on other sites.
- */r/GameOfThrones* A community based around the television adaptation of the fantasy novel series *A Song of Ice and Fire* (Martin 1996), in which members discuss the plot of the show, share pictures and video clips, and post links to external resources.
- */r/videos*, a community for posting links to interesting videos hosted on external sites.

The examples given above are by no means intended to be a complete taxonomy of social media communities; such a taxonomy does not currently exist in the literature, and is an opportunity for further research.

Additionally, the argument is not being made that communities are exclusively created-content or discovered-content. While several communities on Reddit have community guidelines which specifically exclude posts that directly link to other sites (such as */r/GameDev*, a game developer's discussion forum[1]), these are rarer.

Having established this distinction between communities on Reddit, it becomes possible to evaluate the types of manipulation identified over the course of this reserach in the same terms. Given that the purpose of using manipulation techniques is to improve the ranking of content submitted to a social media community, an evaluation of how manipulation techniques relate to the content focus of communities is now warranted.

6.2.2 Understanding the impact of manipulation

Manipulation may now be discussed in the context of the communities in which it is found. Several different types of impact were identified over the course of the research, and these are now discussed in the context of the communities in which they are situated.

6.2.2.1 Manipulation is not always negative

The research presented in this thesis has noted that content ranking manipulation techniques are not inherently detrimental to a community on their own; indeed, manipulation techniques can be indistinguishable from high-quality content, and in some cases have been reported to have a positive impact on user experience. The most striking example of this is the 'long text' type of manipulation, which several participants in the study described in Chapter 5 reported as having a *positive* impact on their brows-

[1] `http://reddit.com/r/gamedev`, active as at October 24 2014

ing experience, due to the post title providing them with more context to decide whether or not they want to read the rest of the post.

This allows parallels to be drawn with search engine optimisation (SEO) techniques, which are a collection of content ranking manipulation techniques targeted towards search engines. SEO techniques involve the use of knowledge of the ranking methods used by search engines to tailor the content of web pages by the author (or, commonly, an agent contracted by the owner), such that the search engine favours that web page in its ranking. SEO techniques are typically divided into "white-hat" techniques, in which the content is designed to be easily interpreted by the search engine's software while at the same time improving the clarity and accessibility of the web page, and "black-hat" techniques, in which the content of the web page is designed to exploit the behaviour of the search engine's software without improving the user experience (Ntoulas, Najork, Manasse and Fetterly 2006). In more concrete terms, white-hat SEO techniques are the equivalent of designing the content so that search engines can gather usefully identifying information from it, to which search queries can be effectively matched, while black-hat SEO techniques are all methods designed to improve ranking without improving the content itself.

Manipulation on social media sites that results in a positive impact upon the users can be considered to be the parallel of white-hat SEO: when a user submits content to a social media site that is accessibile, easy to understand and is readible, that content receives more favourable attention than content that does not. This assertion may be made based on comments from participants presented in Section 5.3, in which users reported multiple-line post titles as improving their experience, by providing more information about the post prior to the user making a decision to click on

the link.

These techniques remain a form of manipulation. However, as has been demonstrated, manipulation is not an innately negative thing; rather, appropriate use of manipulative techniques can improve the overall user experience of social media sites.

6.2.2.2 Manipulation impacts become more severe when more of it appears

A repeated theme among interviewed users was one of manipulation fatigue: when users repeatedly encountered the same kind of manipulation in a social media community, they felt more likely to leave that community.

This issue may be exacerbated by "bandwagon" effects, in which a particular kind of manipulation may increase in prominence over time. These bandwagons can be explained through the following process: a piece of content is submitted to a site, making use of a manipulation technique, and becomes highly ranked; other users notice the high ranking of this content, and submit further content using the same manipulation technique; eventually, the community begins to react against the amount of manipulation.

6.2.2.3 Manipulation can reduce the appeal of content

Participants repeatedly stated that they view Reddit to find new and interesting content. Because Reddit is only one example of sites whose purpose is to act as a community where new and interesting content may be found, The less new and interesting content they find, the more likely they are to drift away - either to a different part of the site, or to another site altogether.

Many kinds of manipulation cause people to ignore content. The more a person ignores content, the less likely they are to remain a member of the community, because they'll find that they're getting less out of the community.

6.2.2.4 Manipulation and its relationship to submitted content

Having established a method of categorising different types of manipulation, a pattern emerges: some manipulation techniques, as identified by this research, have depend upon the details of the content whose ranking is being manipulated, while techniques may be applied to any content regardless of its substance.

For example, the *"reference to self"* type of manipulation, in which the poster makes a reference to themselves in relation to the content (such as an example given in Section 4.4, *"My homemade mermaid costume!"*, is considered a type of manipulation that depends on the content, because the manipulation directly references the material linked to.

Similarly, the *"reference to Reddit"* type of manipulation, in which the poster directly addresses the audience of Reddit (for example, *"Hey Reddit. Check out this dog!"*), does not depend on the specific content being referred to, and can readily be re-applied in a generic sense. For example, one may easily re-purpose the technique by changing the subject: *"Hey Reddit! Check out this news article!"*.)

Having identified this distinction, the two categories are now defined as:

1. types of manipulation that are *dependent* on the content being linked to; and

2. types of manipulation that are *independent* of the content being linked

to.

Using these categories, each of the elaborated types of manipulation identified in Chapter 5 was reviewed, and each type of manipulation classified. The classification, and the justification for the decision made, is shown in Figure 6.1.

Type of Manipulation	Refers to Content	Independent of Content	Justification
Appeal/beg		✓	Begging for attention to be drawn to content does not necessarily rely on the subject itself.
Offensive		✓	Offensive language draws attention on its own
Faking stupidity	✓		Presenting to be ignorant relies on not understanding the linked content
Misleading	✓		Misleading means the content being linked to is different than advertised
Reference to self	✓		"I made X" cannot apply to all kinds of content
Exclamation marks		✓	Exclamation marks draw attention on their own
Asking for feedback	✓		Asking for feedback means people need to provide feedback on the content that is linked
New thread old topic	✓		Creating new threads based on old content means that the new thread must relate to the content
Caps		✓	Large letters draw attention on their own
Reference to Reddit		✓	Phrases such as "Hey Reddit" can be associated with any content
Reference to Relationships		✓	Phrases such as "My girlfriend made this" can be associated with any content
Long text		✓	Extending the length of the post increases the visibility of the text regardless of the content being linked
Link bait		✓	Hyperbolic/teasing text (e.g., "Wow. Just wow.") is enticing on its own and doesn't rely on support from the content.
Advertising	✓		Advertising directly promotes the content itself
Self promotion	✓		"We made this thing" depends directly on the content you're promoting
Implicit brand marketing	✓		The brands are embedded directly in the content
"I'm special"		✓	The identity of the poster is drawing attention, not the content they're posting
Cakeday		✓	The statement "it's my cake day" is presented as justification for up votes on its own
Implicit request for upvotes		✓	Statements like "Why isn't this getting attention" do not directly depend on the content that is not getting attention
Explicit request for upvotes	✓	✓	Explicit requests for up votes make reference to the content or attributes of the content. ("You need more upvotes, this is heartwarming.")
Organising mass votes	✓		Mass vote organisation can be done independently of even knowing what the content is.

Figure 6.1: Types of manipulation, and whether or not they refer to the content being linked to.

Importantly, manipulation does not appear to be significantly biased towards content dependence or independence. There appears to be a reasonably even split between manipulation techniques that refer to the content being posted, and those that are independent of the content being posted. This point is worthy of further discussion, as it means that the scope of discussion for manipulation impact need not be constrained.

6.3 Manipulation and Relevance

The topic of *relevance* was introduced in Section 2.3.7, and describes the various forms in which information can be relevant to a user. Among the various types of relevance described by Saracevic (2007), we find *system relevance*, which describes relevance determined by an algorithm.

The core argument of this thesis has been that manipulation's end goal is to influence the global systemic relevance of content. With this in mind, the most straightforward method of addressing manipulation is to address its impact on the global systemic relevance. The final goal of a manipulation is to increase ranked popularity of content submitted to the site. As a consequence, preventing that manipulation from increasing the score of content defeats this manipulation.

This section, therefore, proposes several techniques to mitigate the impact of manipulation in a social media site.

- Section 6.3.1 discusses the difference between global and per-user approaches to selecting and presenting content on social media sites.
- Section 6.3.2 presents possible methods for addressing attempts to manipulate global systemic relevance systems.

6.3.1 Global and per-user filtering approaches

Reddit is a social media site that, like many social media sites, uses system relevance to determine the prominence of content. Reddit's application of system relevance is *global*: all users who visit a subreddit at the same time see the same content, and in the same order. This is presented in contrast to other social media sites, such as Facebook, whose system relevance is constructed to present different content to different users - that is, two users following the same person will not necessarily see the same content posted by that person.

The Reddit-style system of selecting and ordering content the same way for all users is referred to in this thesis as a *global systemic relevance* approach, while the Facebook-style system of individually customising the selection and ordering of content on a per-user basis is referred to as a *per-user systemic relevance* approach.

This research is not concerned with making a judgement regarding which of these two methods is better, and both have arguments in their favour. Global systemic relevance approaches allow users to discover content that they would not have otherwise searched for, while per-user systemic relevance allows users to not have to spend large amounts of time to find information that they care about (boyd 2008).

This section starts with the assumption that a site that currently uses a global systemic relevance system does not wish to replace it with a different system for selecting and ordering content for users. Because this thesis has concerned itself with a case study of a site that uses a global approach, its conclusions may not necessarily apply to sites that use per-user content selection.

This thesis has demonstrated that global approaches to systemic rel-

evance are vulnerable to the effects of manipulation techniques. These techniques can have negative effects on the population of the site, ranging from inducing mild dissatisfaction to users leaving parts of the site or even departing the site entirely. Manipulation can also have positive effects on the users, if it leads to users finding it easier to locate information that they care about. The research presented in this thesis did not attempt to reach any conclusions about the relative prevalence of negative impacts versus positive impacts; however, the researcher believes that the number of negative impact types identified in Chapter 5 provides reasonable grounds to assume that more manipulation is negative than is positive.

6.3.2 Addressing global systemic manipulation

Given these two assumptions - that sites that use global systemic relevance approaches are unwilling to replace them, and that manipulation can lead to negative impacts - it is now possible to use the lens of manipulation-as-relevance-modification presented in this thesis as a tool for examining how manipulation can be mitigated.

The remainder of this section presents several options for mitigating the impact of manipulation in global systemic relevance. Some of these options are currently in operation in social media sites; where these are known to the researcher, they are discussed, though the study of their efficacy as implemented is beyond the scope of this research.

As part of the discussion of each presented technique, notes on its applicability to created-content and discovered-content communities are also presented.

6.3.2.1 Manual curation

Curation, in the context of content submitted to social media sites, means a process in which content is manually selected by users with moderation access and promoted above other content.

Human-driven curation augments the quality selection process provided by algorithm-driven content selection: regular, non-moderating users submit and vote on content, which allows moderators to more easily identify high-quality content to make selections from. (The astute reader will notice that this means that manipulation remains possible to a degree, due to the fact that source material for moderation is subject to manipulation.) Manual curation allows moderators to observe and correct for any overabundance of content that they consider to be accumulating on the site.

The quality of a human-driven curation system depends on the quality of its human curators. "Quality", for moderators, has multiple influencing factors: the frequency with which moderators identify and promote new content, the compatibility of the moderator's taste with the community's collective taste, and the quality and volume of content available for the moderator to select from. Human-driven curation also increases the amount of work that must be performed, which means that that site owners must either rely on volunteer work or pay for professional moderation.

The term "curation" is distinct from "moderation", in that curation takes an active role in promoting good content, while the primary goal of moderation is to reduce the impact of bad behaviour through the editing and deleting of posts, and by removing disruptive users from a community.

A partial, ad-hoc implementation of this process can be seen in the subreddit */r/bestof*, in which users post direct links to other posts (in other subreddits) that they consider to represent the "best of" Reddit as a whole.

Another example of curation in action is the forum *Something Awful*, which was discussed in Section 2.4.2.2. The Something Awful forum has an attached website[2], which presents selections of the best threads, comments and content submitted by the forum users.

Curation has applications outside of social media. The online video game store Steam[3], while not based around user-generated content, uses manual curation as part of an effort to make it easier to discover new content.

Manual curation is applicable to both discovered-content and created-content communities. In both cases, because a human moderator has the ability to make judgements about whether content is suitable for the community, the final decision of content prominence is no longer entirely driven by the points earned by content.

6.3.2.2 Automatic moderation

Automatic moderation is the use of moderation software to supplement human moderation. Automatic moderation does not replace human moderation, but can reduce the workload of human moderators by automating common tasks. For example, if a social media site has the ability to let users report content, an automatic moderation system can respond to those reports immediately.

An example of automatic moderation is the AutoModerator bot, which "automates straightforward moderation tasks by automatically performing actions based on defined conditions" (Birch 2014). The AutoModerator is configured with a set of rules that determine its behaviour; examples given by the authors include automatic removal of content that is reported

[2] http://www.somethingawful.com, active as at October 24 2014

[3] http://store.steampowered.com, active as at October 24 2014

by community members a certain number of times, or of content that includes certain key words or phrases.

Automatic moderation is an implementation of global systemic relevance. When the rules of an automatic moderation tool affect the prominence of content, either through promotion or removal, an algorithm-driven system is making a judgement about the relevance of that content to the community as a whole. This makes automatic moderation vulnerable to the same manipulation as the underlying voting system: a user who learns or infers the rule-set used by an automatic moderation tool will be able to construct their posts to either avoid being removed or edited, or cause the automatic moderation system to promote their content. This issue is addressed by making the rules of an automatic moderation tool able to be changed by moderators in response to conditions within the site.

Automatic moderation is effective against attempts to influence global systemic moderation on discovered-content sites by identifying and reducing duplicates; on discovered-content sites, it is capable of handling emergent attention-seeking manipulation techniques and reducing their visibility. However, because automatic moderation systems *are* embody a form of systemic relevance themselves, they are unable to prevent manipulation of systemic relevance on their own, and therefore rely upon human guidance to be effective.

6.3.2.3 Voting restriction

A global systemic relevance system is open to manipulation when the users providing it with the information it uses to determine the relevance of content are themselves manipulated. However, it follows that this manipulation of the global systemic relevance system is only possible when those manipulated users have the ability to provide that information in

the first place.

Restricting which users may vote restricts the number of votes that are potentially the result of the voter being affected by a manipulation technique. The conditions that determine whether a user has permission to vote on content are up to the site administrator; possibilities include the age of the account, the number of comments or posts they have made, and the number of votes their comments or posts have received. This technique requires that the user comprehend a more complicated system, but prevents a number of manipulation techniques; for example, restricting voting based on age precents users creating a large number of accounts to vote with (a form of the Sybil attack discussed by Douceur).

This technique is used in the technology question-and-answer site Stack Overflow[4], a site in which users post programming questions and answers, and vote on both questions and answers in a similar manner to Reddit. Users who have only recently signed up do not have permission to vote on content, and must accrue reputation points by asking and answering questions before they are granted permission to vote. On Stack Overflow, this technique goes beyond voting permissions, and extends to abilities that, on other sites, would be considered the domain of site moderators and administrators. For example, users with a high reputation - that is, users who have posted questions or answers that have received a high number of votes - are granted the ability to edit *other* people's posts[5].

A related technique to Stack Overflow's implementation is the system seen on Slashdot (Dice Holdings Inc. 2014). As discussed in Section 2.4.1,

[4] `http://stackoverflow.com`, active as at October 24 2014

[5] A similar ability is available on the Something Awful forums (Something Awful LLC 2014b), in which users may modify other user's avatar or nickname. On these forums, however, the qualifying mechanism is different: if a user wishes to change another's avatar, they must simply pay a fee of five dollars.

users do not ordinarily have the ability to vote on comments. However, eligible users - defined by the site's owners as logged-in users who have not declined to participate, have a positive average score for comments they have posted, and who read an average number articles (Malda 1999) - are granted moderation points, which they may spend to up-vote or down-vote comments. This restriction of who may vote may limit the inclination of users to play to the crowd, or otherwise engage in manipulative techniques.

6.3.2.4 Vote weighting

Another technique for ameliorating the effect of manipulation on global systemic relevance systems is to make the votes of a subset of users have more impact on the ranking system than the rest of the population. This is a combination of the manual curation approach with the voting restriction approach; selected users with higher voting "power" can be considered to be the same as curators, while other users may have their voting abilities restricted or eliminated by having a lower-than-average voting system.

The literature does not present a great deal of examples for these techniques being used on public websites; the closest approximation to this technique is the "shadow-banning" approach used by Reddit (and encountered by the author, as related in Section 4.3.4), in which disruptive users lose the ability for their posts and comments to be seen by others.

6.3.2.5 Indirect attempts to prevent manipulation

The final methods to address manipulation of global systemic relevance systems are indirect, and as such should not be considered in the same category as the other techniques discussed in this section. However, their implied attitude toward users who engage in manipulation is amusing,

and they are therefore included in this section as an exercise in creativity.

Slow-banning and *error-banning* are terms used by Atwood (2011) that refer to techniques that degrade the quality of the user's experience on a website. Slow-banning introduces artificial delays in the web site for specific users, in order to induce frustration, while error-banning introduces false errors or apparent bugs in the site software (such as failing to load critical resources like images, ignoring information sent by the user, or logging the user out for no reason). The intent behind these indirect methods is to create a similar result of shadow-banning - that is, the user is prevented, or at the very least dissuaded, from participating in the community, while ideally not realising that they have been banned.

Both slow-banning and error-banning are implemented in the open-source social media site software Drupal, in the form of the optional "Misery" plugin (Drupal Foundation 2013), which allows site administrators to choose from a variety of different methods of tormenting selected users. These include exploiting bugs in certain versions of the Internet Explorer browser to cause it to crash, randomly re-directing the user to other parts of the site, and sending a blank page to the user. The author has not made an attempt to determine how prevalent this plugin is among sites that use the Drupal software, and is not aware of any research in the literature relating to these techniques. Accordingly, they are an interesting avenue of future study for researchers[6].

Slow-banning and error-banning are not a means by which a site administrator could prevent a user from manipulating a global systemic relevance system; they are designed to frustrate a user and slow them down, but not to reduce their impact upon the site. They are therefore useful for preventing manipulation only when used in conjunction with other tech-

[6]This topic is applicable to both evil researchers, and regular researchers.

niques.

6.4 Conclusions

This chapter has integrated the findings from the three phases of data collection presented in this study. The discussion comprised the following:

- Section 6.2 presented a model for understanding manipulation in the context of different types of social media communities. Different types of manipulation were examined under the lens of communities as collectors of created content and communities as collectors of discovered content. In the process of proposing this model of manipulation, it was noted that not all kinds of manipulation are relevant to both of these types of community. Taking into account the fact that the differentiation between created-content and discovered-content communities mean that manipulation means different things in different contexts, this chapter has recognized that the impact of manipulation depends upon the structure of the community itself.

- Section 6.3 linked the work presented in this thesis on manipulation to existing work on relevance, and concluded that manipulation techniques are at their core attempts to influence the system that determines the systemic relevance for the community. Following this, multiple options for dealing with attempts to modify systemic relevance were presented; by allowing site administrators to constrain their attempts to mitigate the impact of manipulation to methods that counter quantitative systemic manipulation, site administrators are able to use this model to potentially reduce their workload.

The following chapter, Chapter 7, concludes the thesis by reviewing the

key findings and contributions made by this work. The methodologies employed during the research are discussed, along with the limitations and opportunities for future work.

7

Conclusions

This chapter concludes the thesis, by reviewing the findings, contributions, methodologies used, limitations of the research, and opportunities for future research.

7.1 Introduction

The initial premise of this research was that *there is insufficient research into manipulation on social media sites*. This chapter returns to this premise, and evaluates the success of the research presented in this thesis in addressing the problem that the premise poses. In doing so, the process undertaken in conducting the research is reviewed, along with the results obtained, the limitations of the methods used, and the opportunities for further research in this area.

Following a review of past research in Chapter 2, the following observations were made:

- Most research into social media sites concentrates on the social *networking* aspect of social media, rather than the content that is posted: the links between users in a socially-oriented site;

- Almost all research into manipulation of social media sites considered very few, specific kinds of manipulation. The attacks under consideration were mostly easily-quantifiable attacks such as the creation of multiple accounts to manufacture votes (the Sybil attack discussed in Douceur's 2002 work);

- Little to no research was being done into the impact of manipulation on the users of social media sites. Most research simply took the approach that manipulation of any kind was harmful to the community, but did not tend to elaborate further. For example, research focused on how content quality could be determined (see Section 2.3.4), or on specific types of attacks against social media sites (see Section 2.5); no qualitative discussion of manipulation at a higher level was being undertaken.

From these observations, and driven by the original premise of the research, a number of objectives were derived for the research to address. They are as follows:

1. to identify what manipulation comprises, from the perspectives of both administrators and of users;

2. to develop and understanding of the impact of manipulation upon users of social media sites;

3. to propose an empirically grounded model of manipulation and impact in the context of social media sites.

These objectives were designed to both explore the initial premise, as well as to directly address the objectives identified earlier in Section 7.1. These objectives were initially stated earlier in this thesis, in Section 1.2.

In order to achieve this objectives, the three separate phases of study were designed and performed. Each of these three phases directly contributed towards the first and second objectives given above; the third objective was address during the integration of the findings from each of the three phases. This three-phase structure was presented in Section 1.4.

The three phases of study were:

- **Phase 1:** The first phase, interviews with administrators and moderators of Reddit, a social media site that served as the case study for this thesis, was designed to develop an initial understanding of how manipulation is seen by the individuals responsible for the operation of a social media site. Examples of manipulation were identified, and a classification of different types of manipulation was developed.

- **Phase 2:** The second phase, a web-based study undertaken with users of the site, built upon the classification of manipulation types identified in the first phase, and used them to gather perspectives on manipulation from users of the site. In doing so, a novel method of collecting data from users browsing a web site was developed and evaluated.

- **Phase 3:** The third phase involved interviews with participants from the second phase - that is, users of the social media site - and created an understanding of the impact of manipulation upon their use of the site.

This project identified a significant gap in the literature surrounding manipulation on social media sites, and subsequently proposed a model for understanding manipulation and the possible impacts on users.

7.1.1 Chapter structure

The remainder of this chapter discusses in further detail the contributions of this research, the limitations of the methodologies used and how future research may address them, and offers suggestions for further studies. Finally, a parting summary is provided, and the reader is wished a fond farewell.

- Section 7.2 reviews the contributions made by each phase of the research, and discusses the implications of each contribution.
- Section 7.3 discusses the limitations of each phase, and reflects upon how further studies may improve upon each.
- Section 7.4 presents a number of opportunities for further research in this area.
- Section 7.5 concludes the thesis, and offers some parting words.

7.2 Contributions

This section summarises and discusses the contributions made by each phase of study, and of the research as a whole. The contributions from each of the phases are identified as being either *substantive*, *theoretical*, or *methodological*. Following the presentation of each of the contributions, the implications of each type of contribution upon future research are discussed.

- *theoretical:* the identification that manipulation of social media sites attempt to influence systemic relevance for all users of a site by influencing other types of relevance for a subset of users;

- *theoretical:* a definition of manipulation, grounded in the context of earlier work on relevance: *manipulation on social media sites is an attempt to change the global algorithmic relevance of content by changing its individual topical and affective relevance*;
- *methodological:* a novel method of rapidly collecting in-situ data from users through the use of a web browser extension;
- *substantive:* identification that manipulation is believed by site administrators to exist, and exploration of the different types of manipulation;
- *substantive:* exploration and identification of different types of manipulation, grounded in user-identified examples of manipulation;
- *theoretical:* construction of a model of types of manipulation, verified by both administrators and of users;
- *substantive:* the discovery that users of social media sites have a different definition of "spam" to "unsolicited commercial messaging";
- *substantive:* collection of data regarding the impact of manipulation on user's use of social media site;
- *substantive:* exploration and classification of types of impact, ranked by severity.

7.2.1 Implications of these contributions

This research has found that manipulation of social media exists. Both administrators and users emphatically stated that they had observed manipulation while using the social media site Reddit, and that it had impacted their use of the site.

Several different types of manipulation were identified. The majority of these types focused on attempts to gain user attention, though others attempted to gain votes in a more direct manner. This focus on attention-grabbing manipulation is directly relatable to Saracevic (2007)'s identification of "affective" types of manipulation.

Manipulation relates to systemic relevance, and attempts to gain an increase in prominence attempt to influence the topical, affective or cognitive relevance of that content for specific users (as per Saracevic's 1975 and 2007 work). For example, when a user attempts to draw attention to a personal relationship when posting, they are attempting to influence the affective relevance of that content to a subset of users who are likely to respond. If these users then vote the content up, all users on the site are more likely to see it, because the ranking system used by the site makes highly-voted content more visible to everyone.

The contributions of this research have several significant implications for future research.

Implications of Theoretical Contributions The research presented in this thesis links the previously poorly-defined topic of manipulation in social media sites to existing theoretical work on relevance. By establishing that manipulation is directly relatable to systemic relevance, it becomes possible to discuss manipulation within the context of existing models of relevance.

This thesis additionally provides a model of manipulation and its impact, which forms a framework within which further research on manipulation may be conducted. Consequently, these contributions support the objectives by creating, extending, and exploring the validity of the model.

No past research has proposed a similar framework that encompasses

multiple different types of manipulation.

Implications of Substantive Contributions The substantive contributions of the thesis come together to support the objectives of the research by gathering significant information on the various species of manipulation in social media sites, using Reddit as a case study. In doing so, an understanding of what manipulation is has been created, along with an understanding of how manipulation can affect the way people use social media sites. These contributions provided the impetus for the development of the web-based data collection tool for Phase 2, which forms a significant methodological contribution of the thesis.

No past research has provided a definition of manipulation in this manner; while past studies have selected a single type of attack, none have attempted to define manipulative behaviour as a whole.

Implications of Methodological Contributions The methodological contributions of this thesis support the objectives by creating a method through which high-quality, empirically grounded user information can be quickly gathered. Due to the infeasibility of gathering qualitative data from a large number of participants in a reasonable amount of time, the in-situ web browser extension allowed for the rapid, tuneable collection of information.

This method has a number of advantages over traditional diary studies, due to the participants gathering and delivering data to the investigators without ever leaving the context of the situation in which the data is generated.

In addition, by designing the web browser extension to provide opportunities for data collection in a subtle, non-intrusive manner, participants are less likely to abandon the study, and instead provide more data over a

longer period of time.

Finally, by making the data collection method a piece of downloadable software that does not require interaction with the investigators to install, new participants can be added to the study with significantly greater ease; indeed, participants themselves can serve as a conduit for recruiting additional participants.

This approach to "diary-study"-style data collection with downloadable web-browser extensions has not appeared in the literature to date, and is a remarkably flexible data-collection tool for future studies to make use of.

7.2.2 Answering the research questions

As originally stated in Section 1.2, the questions posed when this research began were:

RQ1 What are the most prevalent kinds of manipulations taking place on these social media sites?

RQ2 What impact do these types of manipulations have on the communities?

RQ3 How severe are the different types of manipulations in terms of their impact on a community?

RQ4 What can site owners do to address manipulation?

The research presented in this thesis has answered these questions in the following ways:

- RQ1 has been answered by creating a framework for describing types of manipulations in social media sites. This has allowed for de-

tailed discussion of the subtleties of each different type; for example, the discussion of attention-grabbing manipulation presented in Section 4.4.1 identified a multiplicity of different means of manipulation.

- RQ2 and RQ3 have been answered through the analysis of interviews with Reddit users, presented in Chapter 5. A scale of impact severity has been created, and the author was mildly surprised to note that not all forms of manipulation are considered to be negative; indeed, some, like long post titles, was considered sometimes positive.

- RQ4 has been answered by the realisation that manipulation can be interpreted as attempts to influence systemic relevance. As a result, methods for dealing with manipulation can be better focused; additionally, several concrete options for dealing with the manipulation of systemic relevance systems are presented in Section 6.3.2.

7.3 Limitations

This section discusses the overall success of the research in achieving the objectives set out in Section 1.2. As is the case with all PhD theses, the research was limited to a single invidual performing all study design, data collection, analysis, and discussion. While this is the central feature of PhD researcher, the author feels that it is worthwhile to mention this limitation, as it forms much of the context of other limitations in the research.

Other potential limitations, which are discussed in this section, include:

- The scope of the study: the choice of social media site under consideration.

- The participants of the study: participant recruitment.
- Self-reported data.

Limitations of scope: A single social media site, Reddit, was used as the case study in all three phases of the research. This constraint was felt to be reasonable, due to several factors: *a*) Reddit serves as the platform upon which multiple social media communities are built, which means that a wide variety of different types of communities could be studied that maintained a very large degree of similarities to each other; *b*) the estimated user population is very large, with over 80 million visitors per month recorded in the month of October 2013; *c*) by restricting the focus of the study to a single site, the development of the web-based development tool was able to focus its scope towards the structure of only one site.

Limitations of participants: Recruitment of administrators and moderators was performed directly, through a combination of email and private messaging on Reddit, while recruitment of users for the second phase of the research was performed through a public invitation to participate, and recruitment of interviewees in the third phase was performed through a second public invitation to participate. This limited the number of people who were able to provide data, which may have influenced the results available for analysis. Further studies might use additional methods for recruiting larger numbers of participants.

Limitations of data: all data collected was a qualitative nature, and the research did not attempt to quantify the impact of manipulation. The researcher feels that this was appropriate, given the fact that only some kinds of impacts have a quantifiable impact, and it was felt important that

the research not exclude from consideration those that do not. Additionally, all data presented in this thesis was of a self-reported nature. While every effort was taken to ensure the accuracy of the data, self-reported data can be less reliable and precise than data collected using more objective means. The researcher feels that this risk is acceptable and appropriate to the study, given that it is an accepted aspect of this kind of research (Lazar et al. 2010). However, the design of the research included features intended to ameliorate this potential problem; the web-based data collection study presented in Chapter 4 collected data from users immediately, and did not allow users to ruminate on their observations before the researcher could gather the data from them: the delay between a user observing a data point and the researcher receiving it was on the order of several seconds, as opposed to the several hours, days or weeks commonly seen in other diary studies (Rieman 1993, Czerwinski et al. 2004, Teevan et al. 2004).

As Rieman (1993) notes, participants in a diary study must usually be convinced to make a considerable effort over a period of time to record their activities. By collecting data immediately, participants were spared the effort of managing the data they collected.

7.4 Future Work

This section discusses opportunities for further research into this area. The research presented in this thesis was non-longitudinal in its nature; while the researchers feel that it is unlikely for the nature of ranking manipulation to change, longer-term impacts on communities as a whole may be observed over a longer period of time.

In addition, by creating a framework for the discussion of manipula-

tion in social media, the research has allowed for more precise investigation of different kinds of manipulation techniques.

Detecting trends in manipulation techniques: Further research is warranted into criteria and mechanisms for detecting when a trend of manipulation is noticed in a social media community. Several participants of Phase 2 (discussed in Chapter 4) made it clear to the researcher that, while they did not particularly mind the occasional occurence of manipulative content, witnessing a social media community fill with a certain kind of manipulation caused them to leave the community entirely. The deleterious nature of this kind of impact on users has severe potential consequences for the community as a whole. It would therefore be interesting to investigate whether trends of manipulation may be detected before any potential effects become too much of a problem.

Manipulation from moderators: This thesis only investigated manipulation in the form of posts and comments submitted by users of social media sites, and did not consider another potential source of manipulation: administrators and moderators themselves. Administrators and moderators have a high degree of control over the community, ranging from the ability to ban users to making changes to the design and presentation of the site. During interviews with moderators in Phase 1, a few administrators made oblique references to other moderators exhibiting what they termed 'power-hungry' behaviour, including banning users whom the participants did not feel had warranted being banned.

Separately, towards the end of the research, the researcher noticed a case in which the creator of a subreddit had manipulated the presentation of their site to prepend a numeral to the display of the number of users who had subscribed to the community, with the result that the subreddit

appeared to have a significantly higher number of subscribers than it actually did. When the moderators of the community were asked about this practice, they defended it as a technique to *"give people a morale boost"*:

> *"Again, increasing the number had an immediate effect on subscription rate and user activity. This was done knowing that people would find out eventually and possibly get angry. Nobody is trying to hide anything, it's just important to give people a morale boost early on.* [1] *"*

This comment was posted in December 2013, when the subreddit had approximately 1,400 subscribers; ten months later, in October 2014, the subreddit had approximately 5,500 subscribers.

The exploration of moderator-driven manipulation opens up additional avenues for studying manipulation in social media sites, and would be very interesting to research further.

Quantifying impacts by correlating presence of manipulation against subscriber count and other factors: An exploration into how different kinds of manipulation affect quantifiable factors in how social media sites are used is warranted. While the mechanics of attacks like the Sybil attack are well-studied, the longer-term impact on the number of people participating in the community is unknown.

Further theory development: Conceptual models are well-established in academia, but can suffer from the problem of limited accessibility to non-academics. Further development of the model proposed in this thesis

[1] http://www.reddit.com/r/gamedev/comments/1svxud/announcing_runity2d_a_subreddit_for_2d_and_25d/ce20fsv; an image gallery showing the manipulation technique in more detail is available at http://imgur.com/a/3ZhzP.

would render it more accessible to administrators, users of social media sites, tool developers, and other people who interact with social media sites.

Exploration of moderator feelings on manipulation: This research did not attempt to determine *why* moderators believed that manipulation was a problem; rather, it attempted to examine the manipulation itself, and not the underlying reasoning for why moderators believed it to be a problem. The discussion of the impact of manipulation, presented in Section 5.3, discusses user perspectives on how manipulation impacts them, but they may be different to how moderators feel about the same topic.

Further use of web browser tool: The web-based data collection tool used in Phase 2 of this research has significant scope for use beyond this study. As has already been demonstrated in Chapter 4, in-situ data collection on the web provides fast, flexible, high-quality data collection that does not interfere with the user's natural browsing behaviour. Applying the same methodology to the study of other social media sites is an obvious extension of the technique, but more elaborate opportunities exist; it is reasonable to expect that the data-collection technique could be used to gather information on a wide variety of sites.

7.5 Parting Words

This research began with the premise that *there is insufficient research into manipulation on social media sites*. Over the course of the research, this thesis has addressed this lack of research by identifying what manipulation is, how adminstrators, moderators and users perceive it, and what impacts it has upon the users of the site. Future research is now able to investi-

gate manipulation still further, and, it is hoped, continue to improve the experience of social media and content discovery for both users, and of administrators and moderators.

References

4chan, LLC: 2014, 4chan. [Accessed 23/10/2014].
URL: *http://www.4chan.org/*

Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G.: 2008, Finding high-quality content in social media, *the international conference*, ACM Press, New York, New York, USA, pp. 183–194.
URL: *http://doi.acm.org/10.1145/1341531.1341557*

Althoff, T., Danescu-Niculescu-Mizil, C. and Jurafsky, D.: 2014, How to Ask for a Favor: A Case Study on the Success of Altruistic Requests.
URL: *http://arxiv.org/abs/1405.3282*

Anderson, A., Huttenlocher, D., Kleinberg, J. and Leskovec, J.: 2012a, Discovering value from community activity on focused question answering sites: a case study of stack overflow, *KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM Request Permissions.
URL: *http://portal.acm.org/citation.cfm?id=2339530.2339665*

Anderson, A., Huttenlocher, D., Kleinberg, J. and Leskovec, J.: 2012b, Effects of user similarity in social media, pp. 703–712.
URL: *http://dl.acm.org/citation.cfm?id=2124378*

Anonymous SRSTrackerBot author: 2013, I am the SRStracker bot, source code inside. [Accessed 20/10/2014].
URL: *http://www.reddit.com/r/self/comments/18sf1e/i_am_the_srstracker_bot_source_code_inside/*

Apple Inc.: 2014, Safari. [Accessed 23/10/2014].
URL: *https://www.apple.com/safari/*

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A. and Stoica, I.: 2009, Above the Clouds: A Berkeley View of Cloud Computing, *Communications of the ACM* **53**(4), 50–58.

Atwood, J.: 2011, Suspension, Ban or Hellban? [Accessed 22/10/2014].
URL: *http://blog.codinghorror.com/suspension-ban-or-hellban/*

Baldonado, M. Q. W. and Winograd, T.: 1997, SenseMaker, *the SIGCHI conference*, ACM Press, New York, New York, USA, pp. 11–18.
URL: *http://portal.acm.org/citation.cfm?doid=258549.258563*

Bates, M. J.: 1989, The design of browsing and berrypicking techniques for the online search interface, *Online Information Review* **13**(5), 407–424.
URL: *http://www.emeraldinsight.com/10.1108/eb024320*

Baym, N. K.: 2000, *Tune in, log on: Soaps, fandom, and online community*, Vol. 3, Sage.

Bernstein, M., Monroy-Hernández, A., Harry, D., André, P., Panovich, K. and Vargas, G.: 2011, 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community, *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pp. 50–57.

Bethesda Softworks LLC: 2014, The Elder Scrolls Official Site — Skyrim. [Accessed 20/10/2014].
URL: *http://www.elderscrolls.com/skyrim*

Birch, C.: 2013, Subreddits by subscribers - stattit. [Accessed 04/09/2013].
URL: *http://stattit.com/subreddits/by_subscribers/*

Birch, C.: 2014, Deimos/AutoModerator. [Accessed 10/10/2014].
URL: *https://github.com/Deimos/AutoModerator*

Borlund, P.: 2003, The concept of relevance in IR, *Journal of the American Society for Information Science* **54**(10), 913–925.
URL: *http://doi.wiley.com/10.1002/asi.10286*

Boyatzis, R. E.: 1998, *Transforming qualitative information: Thematic analysis and code development*, Sage.

boyd, d.: 2008, Facebook's Privacy Trainwreck, *Convergence: The International Journal of Research into New Media Technologies* **14**(1), 13–20.

Braun, V. and Clarke, V.: 2006, Using thematic analysis in psychology, *Qualitative research in psychology* **3**(2), 77–101.

Brunton, F.: 2013, *Spam: A Shadow History of the Internet*, Infrastructures series, MIT Press.
URL: *http://books.google.com.au/books?id=QF7EjCRg5CIC*

Burrell, G. and Morgan, G.: 1979, *Sociological paradigms and organisational analysis*, Vol. 248, London: Heinemann.

Bury, R., Deller, R., Greenwood, A. and Jones, B.: 2013, From Usenet to Tumblr: The changing role of social media, *Participations: journal of audience and reception studies* **10**(1), 299–318.

Butler, B., Sproull, L., Kiesler, S. and Kraut, R.: 2007, Community effort in online groups: Who does the work and why?, *Human-Computer Interaction Institute, CMU* p. 90.

Buttfield-Addison, P., Manning, J. and Nugent, T.: 2014, *Learning Cocoa with Objective-C: Developing for the Mac and IOS App Stores*, O'Reilly Media.
URL: *http://books.google.com.au/books?id=Y5biAgAAQBAJ*

Charmaz, K.: 2006, *Constructing grounded theory: A practical guide through qualitative research*, London: Sage.

Chi, E. H., Pirolli, P., Chen, K. and Pitkow, J.: 2001, Using Information Scent to Model User Information Needs and Actions and the Web, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, pp. 490–497.
URL: *http://doi.acm.org/10.1145/365024.365325*

Cho, C. H., Martens, M. L., Kim, H. and Rodrigue, M.: 2011, Astroturfing global warming: It isn't always greener on the other side of the fence, *Journal of business ethics* **104**(4), 571–587.

Corbin, J. M. and Strauss, A.: 1990, *Basics of qualitative research: Grounded theory procedures and techniques.*, Sage Publications, Inc.

Cosijn, E. and Ingwersen, P.: 2014, Dimensions of relevance, *Information Processing & Management* **36**(4), 533–550.
URL: *http://www.sciencedirect.com/science/article/pii/S0306457399000722*

Crabtree, B. F. and Miller, W. (eds): 1999, *Doing Qualitative Research*, Sage Publications.

Craigslist: 2014, craigslist. [Accessed 23/10/2014].
URL: *http://www.craigslist.org*

Czerwinski, M., Horvitz, E. and Wilhite, S.: 2004, A diary study of task switching and interruptions, *the 2004 conference*, ACM Press, New York, New York, USA, pp. 175–182.
URL: *http://portal.acm.org/citation.cfm?doid=985692.985715*

Daft, R. L. and Lengel, R. H.: 1986, Organizational Information Requirements, Media Richness and Structural Design, *Management Science*

32(5), 554–571.
URL: *http://pubsonline.informs.org/doi/abs/10.1287/mnsc.32.5.554*

Dawkins, R.: 1989, *The Selfish Gene*, Oxford ; New York : Oxford University Press, New York.

Debevec, K. and Romeo, J. B.: 1992, Self-Referent Processing in Perceptions of Verbal and Visual Commercial Information, *Journal of Consumer Psychology* .
URL: *http://www.sciencedirect.com/science/article/pii/S1057740808800460*

Diakopoulos, N. and Naaman, M.: 2011, Towards quality discourse in online news comments, *the ACM 2011 conference*, ACM Press, New York, New York, USA, pp. 133–142.
URL: *http://portal.acm.org/citation.cfm?doid=1958824.1958844*

Dice Holdings Inc.: 2014, Slashdot: News for nerds, stuff that matters. [Accessed 11/10/2014].
URL: *http://slashdot.org/*

Dick, B.: 2005, Grounded theory: a thumbnail sketch.

Dieberger, A., Dourish, P., Höök, K., Resnick, P. and Wexelblat, A.: 2000, Social navigation: techniques for building more usable systems, *interactions* **7**(6), 36–45.
URL: *http://dl.acm.org/citation.cfm?id=352587*

Douceur, J. R.: 2002, The Sybil Attack, *Revised Papers from the First International Workshop on Peer-to-Peer Systems*, Springer-Verlag, London, UK, pp. 251–260.
URL: *http://dl.acm.org/citation.cfm?id=646334.687813*

Drever, E.: 1995, *Using Semi-Structured Interviews in Small-Scale Research. A Teacher's Guide.*, ERIC.

Drupal Foundation: 2013, Misery — Drupal.org. [Accessed 22/10/2014].
URL: *https://www.drupal.org/project/misery*

Electronic Arts Inc.: 2014, The Sims - Homepage - Official Site. [Accessed 20/10/2014].
URL: *http://www.thesims.com/*

Eslami, M., Aleyasen, A., Karahalios, K., Hamilton, K. and Sandvig, C.: 2015, FeedVis: A Path for Exploring News Feed Curation Algorithms, *the 18th ACM Conference Companion*, ACM Press, New York, New York, USA, pp. 65–68.
URL: *http://dl.acm.org/citation.cfm?doid=2685553.2702690*

European Computer Manufacturers Association: 2011, ECMAScript Language Specification, *Technical report*.

Facebook Inc.: 2013, Mark Zuckerberg: Is Connectivity a Human Right? - Facebook Newsroom. [Accessed 29/08/2013].
URL: *http://newsroom.fb.com/News/693/Mark-Zuckerberg-Is-Connectivity-a-Human-Right*

Fisher, D., Smith, M. and Welser, H. T.: 2006, You Are Who You Talk To: Detecting Roles in Usenet Newsgroups, *System Sciences, 2006. HICSS '06. Proceedings of the 39th Annual Hawaii International Conference on*, p. 59.
URL: *http://portal.acm.org/citation.cfm?id=1109711.1109748*

Garcia, D., Mendez, F., Serdült, U. and Schweitzer, F.: 2012, Political polarization and popularity in online participatory media: an integrated approach, *PLEAD '12: Proceedings of the first edition workshop on Politics,*

elections and data, ACM Request Permissions.
URL: *http://dl.acm.org/citation.cfm?id=2389665*

Gilbert, E.: 2013, Widespread underprovision on Reddit, *CSCW '13: Proceedings of the 2013 conference on Computer supported cooperative work*, ACM Request Permissions.
URL: *http://dl.acm.org/ft_gateway.cfm?id=2441866&ftid=1347777*

Gillespie, T.: 2014, The Relevance of Algorithms, *Media technologies: Essays on communication, materiality, and society*, MIT Press, p. 167.

Glaser, B. G.: 1992, *Emergence vs forcing: Basics of grounded theory analysis*, Sociology Press.

Glaser, B. G. and Strauss, A. L.: 1967, The discovery of grounded theory: Strategies for qualitative research, *Technical report*, Chicago.

Goffman, E.: 1959, *The Presentation of Self in Everyday Life*, Garden City, NY Double Day.

Goldstein, N., Manning, J. and Buttfield-Addison, P.: 2010, iPhone and iPad Game Development For Dummies, *iPhone and iPad Game Development For Dummies* .
URL: *http://portal.acm.org/citation.cfm?id=1951911*

Google Inc.: 2014, Google Chrome. [Accessed 23/10/2014].
URL: *http://www.google.com/chrome/*

Grinberg, M.: 2014, *Flask Web Development: Developing Web Applications with Python*, O'Reilly Media, Inc.

Guba, E. G. and Lincoln, Y. S.: 1994, Competing paradigms in qualitative research, *Handbook of qualitative research* **2**, 163–194.

Guo, L., Tan, E., Chen, S., Zhang, X. and Zhao, Y. E.: 2009, Analyzing patterns of user content generation in online social networks, *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM Request Permissions.
URL: *http://portal.acm.org/citation.cfm?id=1557019.1557064*

Harper, F. M., Raban, D., Rafaeli, S. and Konstan, J. A.: 2008, Predictors of answer quality in online Q&A sites, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, pp. 865–874.
URL: *http://doi.acm.org/10.1145/1357054.1357191*

Harrison, S. and Dourish, P.: 1996, Re-place-ing space: the roles of place and space in collaborative systems, pp. 67–76.
URL: *http://dl.acm.org/citation.cfm?id=240193*

Herodotus: 430 BCE, *On The Customs of the Persians*, fordham.edu.
URL: *http://www.fordham.edu/halsall/ancient/herodotus-persians.asp*

Heroku Inc.: 2014, Heroku — Cloud Application Platform. [Accessed 23/10/2014].
URL: *https://www.heroku.com/*

Hill, W. C., Hollan, J. D., Wroblewski, D. and McCandless, T.: 1992, Edit wear and read wear, *the SIGCHI conference*, ACM Press, New York, New York, USA, pp. 3–9.
URL: *http://portal.acm.org/citation.cfm?doid=142750.142751*

Hillery, G. A.: 1955, Definitions of community: Areas of agreement, *Rural sociology* **20**, 111–123.

Hiltz, S. R.: 1985, *Online Communities: A Case Study of the Office of the Future*, Vol. 2 of *Human/computer interaction*, Intellect Books.
URL: *http://books.google.com.au/books?id=iDDfdQjxEL0C*

Hogg, T. and Lerman, K.: 2012, Social Dynamics of Digg, *arXiv.org* .
URL: *http://arxiv.org/abs/1202.0031v1*

Imgur Inc.: 2009, Imgur - The Simple Image Sharer. [Accessed 23/10/2014].
URL: *http://imgur.com/*

Indratmo: 2010, *Supporting exploratory browsing with visualization of social interaction history*, PhD thesis, University of Saskatchewan.
URL: *http://ecommons.usask.ca/handle/10388/etd-01292010-140030*

Indratmo and Vassileva, J.: 2009, Social Interaction History: A Framework for Supporting Exploration of Social Information Spaces, *2009 International Conference on Computational Science and Engineering*, IEEE, pp. 538–545.
URL: *http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5284250*

Internet Archive: 2012, The Internet Archive: Digg.com as at January 1, 2012. [Accessed 23/10/2014].
URL: *https://web.archive.org/web/20120101011109/http://digg.com/*

IPS Inc.: 2013, IPS Community Suite. [Accessed 23/10/2014].
URL: *http://www.invisionpower.com/*

John, J. P., Yu, F., Xie, Y., Krishnamurthy, A. and Abadi, M.: 2011, deSEO: combating search-result poisoning, *SEC'11: Proceedings of the 20th USENIX conference on Security*, USENIX Association.
URL: *http://portal.acm.org/citation.cfm?id=2028067.2028087*

Kaltenbrunner, A., Gomez, R. and Lopez, V.: 2007, Description and Prediction of Slashdot Activity, *Fifth Latin American Web Congress*, IEEE, pp. 57–66.
URL: *http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=4383159*

Kaplan, A. M. and Haenlein, M.: 2010, Users of the world, unite! The challenges and opportunities of Social Media, *Business Horizons* **53**(1), 59–68.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0007681309001232*

Kaptelinin, V.: 2003, UMEA: Translating Interaction Histories into Project Contexts, *the conference*, ACM Press, New York, New York, USA, pp. 353–360.
URL: *http://portal.acm.org/citation.cfm?doid=642611.642673*

Kittur, A. and Kraut, R. E.: 2008, Harnessing the wisdom of crowds in wikipedia: quality through coordination, *CSCW '08: Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM Request Permissions.
URL: *http://portal.acm.org/citation.cfm?id=1460563.1460572*

Kjeldskov, J. and Graham, C.: 2003, A review of mobile HCI research methods, *Human-computer interaction with mobile devices and services* pp. 317–335.
URL: *http://www.springerlink.com/index/4NJMXBLUKXFLJ9TM.pdf*

Know Your Meme: 2014, Bump — Know Your Meme. [Accessed 20/08/2014].
URL: *http://knowyourmeme.com/memes/bump*

Komito, L.: 1998, The Net as a foraging society: Flexible communities, *The information society* **14**(2), 97–106.

Konstan, J. A. and Chen, Y.: 2007, Online field experiments: Lessons from CommunityLab, *Proceedings of Third International Conference on e-Social Science*, Citeseer.

Kotz, D.: 2013, Number injured in marathon bombing revised downward to 264, *Boston Globe* .
URL: *http://www.bostonglobe.com/lifestyle/health-wellness/2013/04/23/number-injured-marathon-bombing-revised-downward/NRpaz5mmvGquP7KMA6XsIK/story.html*

Kumar, R., Novak, J. and Tomkins, A.: 2006, Structure and evolution of online social networks, *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM Request Permissions.
URL: *http://portal.acm.org/citation.cfm?id=1150402.1150476*

Lambert, B.: 2007, As Prostitutes Turn to Craigslist, Law Takes Notice, *nytimes.com* .
URL: *http://www.nytimes.com/2007/09/05/nyregion/05craigslist.html*

Lampe, C., Johnston, E. and Resnick, P.: 2007, Follow The Reader: Filtering Comments on Slashdot, *CHI*, ACM Press, New York, New York, USA, p. 1253.
URL: *http://portal.acm.org/citation.cfm?doid=1240624.1240815*

Lazar, J., Feng, J. H. and Hochheiser, H.: 2010, *Research methods in human-computer interaction*, John Wiley & Sons.

Lengel, R. H. and Daft, R. L.: 1988, The Selection of Communication Media as an Executive Skill., *Academy of Management Executive* **2**(3), 225–232.
URL: *http://connection.ebscohost.com/an/4277259*

Lewis, C.: 2014, Temporal Dark Patterns, *Irresistible Apps*, Apress, pp. 103–110.
URL: *http://link.springer.com.ezproxy.utas.edu.au/chapter/10.1007/978-1-4302-6422-4_9*

Linden, G., Smith, B. and York, J.: 2003, Amazon. com recommendations: Item-to-item collaborative filtering, *Internet Computing, IEEE* **7**(1), 76–80.

Louise Barriball, K. and While, A.: 1994, Collecting Data using a semi-structured interview: a discussion paper, *Journal of advanced nursing* **19**(2), 328–335.

Lueg, C. and Fisher, D.: 2003, *From Usenet to CoWebs: Interacting with Social Information Spaces*, Computer Supported Cooperative Work, Springer London.
URL: *http://books.google.com.au/books?id=croZMqc0l_oC*

Malda, R.: 1999, Slashdot Moderation. [Accessed 20/08/2014].
URL: *http://slashdot.org/moderation.shtml*

Manning, J. and Buttfield-Addison, P.: 2014, *iOS Game Development Cookbook*, O'Reilly Media.
URL: *http://books.google.com.au/books?id=sohUAwAAQBAJ*

Marchionini, G.: 2006, Exploratory search, *Communications of the ACM* **49**(4), 41.
URL: *http://portal.acm.org/citation.cfm?doid=1121949.1121979*

Martin, G. R. R.: 1996, *A Game Of Thrones*, Random House LLC.

May, T.: 2011, *Social research: Issues, methods and research*, McGraw-Hill International.

Mell, P. and Grance, T.: 2011, The NIST definition of cloud computing.

Metafilter Network Inc.: 2013a, MetaFilter. [Accessed 29/08/2013].
URL: *http://www.metafilter.com/*

Metafilter Network Inc.: 2013b, MetaFilter FAQ. [Accessed 29/08/2013].
URL: *http://faq.metafilter.com/*

Metareddit: 2014, metareddit - all about reddit. [Accessed 08/05/2014].
URL: *http://metareddit.com/*

Mezei, C.: 2006, The Digg Algorithm – Unofficial FAQ. [Accessed 11/10/2014].
URL: *http://www.seopedia.org/tips-tricks/social-media/the-digg-algorithm-unofficial-faq/*

Microsoft Corp.: 2014, Microsoft Internet Explorer. [Accessed 23/10/2014].
URL: *http://windows.microsoft.com/en-us/internet-explorer/download-ie*

Mirkovic, J., Dietrich, S., Dittrich, D. and Reiher, P.: 2004, *Internet Denial of Service: Attack and Defense Mechanisms (Radia Perlman Computer Networking and Security)*, Prentice Hall PTR, Upper Saddle River, NJ, USA.

Morrison, C.: 2014, Random Acts of Pizza - Reddit.com/r/random_acts_of_pizza. [Accessed 21/08/2014].
URL: *http://randomactsofpizza.com/*

Motoyama, M., McCoy, D., Levchenko, K., Savage, S. and Voelker, G. M.: 2011, Dirty jobs: the role of freelance labor in web service abuse, *SEC'11: Proceedings of the 20th USENIX conference on Security*, USENIX Association.
URL: *http://portal.acm.org/citation.cfm?id=2028067.2028081*

Moynihan, S.: 2012, Twitter / ShaunMoynihan: An SEO expert walks into a ... [Accessed 20/02/2013].
URL: *https://twitter.com/ShaunMoynihan/status/273169951802679296*

Mozilla Corp.: 2014, Firefox. [Accessed 23/10/2014].
URL: *https://www.mozilla.org/en-US/firefox/new/*

Muller, M. J. and Kogan, S.: 2010, Grounded theory method in HCI and CSCW, *Cambridge: IBM Center for Social Software* pp. 1–46.

News.me Inc.: 2012, Digg - What the Internet is talking about right now. [Accessed 21/08/2014].
URL: *http://digg.com/*

News.me Inc.: 2014, About — Digg.com. [Accessed 11/10/2014].
URL: *http://digg.com/about*

Nintendo Inc.: 2014, The Official Pokémon Website — Pokemon.com — Explore the World of Pokémon. [Accessed 20/10/2014].
URL: *http://www.pokemon.com/us/*

Ntoulas, A., Najork, M., Manasse, M. and Fetterly, D.: 2006, Detecting spam web pages through content analysis, *Proceedings of the 15th international conference on World Wide Web*, ACM, pp. 83–92.

O'Day, V. L. and Jeffries, R.: 1993, Orienteering in an information landscape: how information seekers get from here to there, *INTERCHI '93*, ACM, ACM Press, New York, New York, USA, pp. 438–445.
URL: *http://portal.acm.org/citation.cfm?doid=169059.169365*

Olson, R. S. and Neal, Z. P.: 2013, Navigating the massive world of reddit: Using backbone networks to map user interests in social media,

arXiv.org .
URL: *http://arxiv.org/abs/1312.3387*

Opera Software ASA: 2014, Opera browser - The alternative web browser - Download free. [Accessed 23/10/2014].
URL: *http://www.opera.com/*

Orlikowski, W. J. and Baroudi, J. J.: 1991, Studying information technology in organizations: Research approaches and assumptions, *Information systems research* **2**(1), 1–28.

phpBB Team: 2013, phpBB. [Accessed 23/10/2014].
URL: *https://www.phpbb.com/*

Pirolli, P.: 1997, Computational models of information scent-following in a very large browsable text collection, *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, ACM, pp. 3–10.

Pirolli, P. and Card, S.: 1995, Information Foraging in Information Access Environments, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, pp. 51–58.
URL: *http://dx.doi.org/10.1145/223904.223911*

Pirolli, P. and Card, S.: 1999, Information foraging., *Psychological review* **106**(4), 643.

Porter, C. E.: 2006, A Typology of Virtual Communities: A Multi-Disciplinary Foundation for Future Research, *Journal of Computer-Mediated Communication* **10**(1), 00–00.
URL: *http://doi.wiley.com/10.1111/j.1083-6101.2004.tb00228.x*

PostgreSQL Global Development Group: 1996, PostgreSQL: The world's most advanced open source database. [Accessed 23/10/2014].
URL: *http://www.postgresql.org/*

Preece, J. and Shneiderman, B.: 2009, The reader-to-leader framework: Motivating technology-mediated social participation, *AIS Transactions on Human-Computer Interaction* **1**(1), 13–32.
URL: *http://www.cs.umd.edu/ ben/papers/Jennifer2009Reader.pdf*

Python Software Foundation: 2001, Python.org. [Accessed 23/10/2014].
URL: *https://www.python.org/*

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A. and Menczer, F.: 2011, Detecting and tracking political abuse in social media, *Proc. of ICWSM* .
URL: *http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2850/3274*

Reddit Inc.: 2012, GitHub: Reddit source code ('Hot' algorithm). [Accessed 05/03/2013].
URL: *https://github.com/reddit/reddit/blob/master/r2/r2/lib/db/_sorts.pyx#L45*

Reddit Inc.: 2013, Reddit FAQ. [Accessed 05/03/2013].
URL: *http://www.reddit.com/wiki/faq*

Reddit Inc.: 2014a, reddiquette - reddit.com. [Accessed 20/10/2014].
URL: *http://www.reddit.com/wiki/reddiquette*

Reddit Inc.: 2014b, Reddit. [Accessed 29/08/2013].
URL: *http://reddit.com*

Reddit Inc.: 2014c, reddit.com: about reddit. [Accessed 11/10/2014].
URL: *http://www.reddit.com/about/*

Rheingold, H.: 1993, The virtual community: Homesteading on the electronic frontier.
URL: *http://books.google.com/books?id=fr8bdUDisqAC*

Rieman, J.: 1993, The diary study: a workplace-oriented research tool to guide laboratory efforts, pp. 321–326.
URL: *http://dl.acm.org/citation.cfm?id=169255*

Robson, C.: 2002, *Real world research: a resource for social scientists and practitioner- researchers*, Oxford.

Salihefendic, A.: 2010, How Reddit ranking algorithms work. [Accessed 29/08/2013].
URL: *http://amix.dk/blog/post/19588*

Saracevic, T.: 1975, RELEVANCE: A review of and a framework for the thinking on the notion in information science, *Journal of the American Society for Information Science* **26**(6), 321–343.
URL: *http://doi.wiley.com/10.1002/asi.4630260604*

Saracevic, T.: 2007, Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance, *Journal of the American Society for Information Science* **58**(13), 1915–1933.
URL: *http://doi.wiley.com/10.1002/asi.20682*

Shit Reddit Says Moderators: 2014, Shit Reddit Says. [Accessed 20/10/2014].
URL: *http://www.reddit.com/r/ShitRedditSays*

Shneiderman, B., Preece, J. and Pirolli, P.: 2011, Realizing the value of social media requires innovative computing research, *Communications of*

the ACM **54**(9), 34.
URL: *http://dl.acm.org/citation.cfm?doid=1995376.1995389*

Short, J., Williams, E. and Christie, B.: 1976, *The Social Psychology of Telecommunications*, John Wiley and Sons Ltd.

Shut Up and Take My Money moderators: 2014, Shut Up And Take My Money. [Accessed 11/07/2013].
URL: *http://www.reddit.com/r/shutupandtakemymoney*

Singh, A. K. and Potdar, V.: 2009, Blocking online advertising - A state of the art, *Industrial Technology, 2009. ICIT 2009. IEEE International Conference on*, pp. 1–10.
URL: *http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4939739*

Smith, D.: 2011, Anonymous Coward - WikiWikiWeb. [Accessed 01/03/2013].
URL: *http://c2.com/cgi/wiki?AnonymousCoward*

Smith, M. A. and Kollock, P.: 1999, *Communities in Cyberspace*, Routledge.
URL: *http://books.google.com.au/books?id=harO_jeoyUwC*

Sobel, S.: 2014, Reddit Enhancement Suite. [Accessed 10/03/2013].
URL: *http://redditenhancementsuite.com/features.html*

Something Awful LLC: 2014a, Comedy Goldmine - The Something Awful Forums. [Accessed 21/08/2014].
URL: *http://forums.somethingawful.com/forumdisplay.php?forumid=21*

Something Awful LLC: 2014b, The Something Awful Forums. [Accessed 21/08/2014].
URL: *http://forums.somethingawful.com/*

SoundCloud Ltd.: 2014, SoundCloud. [Accessed 23/10/2014].
URL: *https://soundcloud.com/*

Spamhaus Project: 2014, The Spamhaus Project - The Definition of Spam. [Accessed 23/10/2014].
URL: *http://www.spamhaus.org/consumer/definition/*

Squad Inc.: 2014, Kerbal Space Program. [Accessed 20/10/2014].
URL: *https://kerbalspaceprogram.com/*

Stanglin, D.: 2013, Student wrongly tied to Boston bombings found dead, *USA Today* .
URL: *http://www.usatoday.com/story/news/2013/04/25/boston-bombing-social-media-student-brown-university-reddit/2112309/*

Star, S. L.: 2007, Living grounded theory: Cognitive and emotional forms of pragmatism, *Sage, London* pp. 75–94.

StatCounter Inc.: 2013, Top 5 Desktop, Tablet & Console Browsers on Oct 2014 — StatCounter Global Stats. [Accessed 20/10/2014].
URL: *http://gs.statcounter.com/#desktop-browser-ww-monthly-201310-201310-bar*

Subreddits.org: 2014, subreddits. [Accessed 21/04/2014].
URL: *http://subreddits.org/*

Suddaby, R.: 2006, From the editors: What grounded theory is not, *Academy of management journal* **49**(4), 633–642.

Svensson, M.: 2000, *Defining and designing social navigation*, PhD thesis, University of Stockholm.

Szabo, G. and Huberman, B. A.: 2010, Predicting the popularity of online content, *Communications of the ACM* **53**(8), 80–88.
URL: *http://dl.acm.org/citation.cfm?id=1787254*

Teevan, J., Alvarado, C., Ackerman, M. S. and Karger, D. R.: 2004, The perfect search engine is not enough: a study of orienteering behavior in directed search, *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, pp. 415–422.

Tufekci, Z.: 2015, Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency , *Colorado Technology Law Journal* **13**.
URL: *http://heinonline.org/HOL/LandingPage?handle=hein.journals/jtelhtel13&div=18&i*

Van Alstyne, M. and Brynjolfsson, E.: 2005, Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities, *Management Science* **51**(6), 851–868.
URL: *http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1050.0363*

vBulletin Solutions: 2014, vBulletin. [Accessed 23/10/2014].
URL: *https://www.vbulletin.com/*

Vickery, G. and Wunsch-Vincent, S.: 2007, *Participative Web And User-Created Content: Web 2.0 Wikis and Social Networking*, Organization for Economic Cooperation and Development (OECD), Paris, France.
URL: *http://dl.acm.org/citation.cfm?id=1554640*

Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M. and Haythornthwaite, C.: 1996, Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual Community, *Annual Review of Sociology* **22**, pp. 213–238.
URL: *http://www.jstor.org/stable/2083430*

Weninger, T., Zhu, X. A. and Han, J.: 2013, An Exploration of Discussion Threads in Social News Sites: A Case Study of the Reddit Community, *WWW '13*, Rio de Janerio, Brazil.
URL: *http://dmserv3.cs.illinois.edu/reddit/paper.pdf*

White, R. W., Kules, B. and Bederson, B.: 2005, Exploratory search interfaces, *ACM SIGIR Forum* **39**(2), 52.
URL: *http://portal.acm.org/citation.cfm?doid=1113343.1113356*

Wilford, J. N.: 1969, Men Walk On Moon, *New York Times* **118**, 1.
URL: *http://www.nytimes.com/learning/general/onthisday/big/0720.html*

Wu, F. and Huberman, B. A.: 2007, Novelty and Collective Attention, *arXiv.org* .
URL: *http://arxiv.org/abs/0704.1158v1*

Y Combinator Inc.: 2014, Hacker News. [Accessed 29/08/2013].
URL: *https://news.ycombinator.com/*

Yahoo! Inc: 2005, Yahoo! Answers. [Accessed 16/02/2013].
URL: *http://answers.yahoo.com*

Yin, R. K.: 2003, *Case Study Research: Design and Methods*, Applied Social Research Methods, SAGE Publications.
URL: *http://books.google.com.au/books?id=BWea_9ZGQMwC*

YouTube LLC: 2014, YouTube. [Accessed 23/10/2014].
URL: *https://www.youtube.com/*

Zhu, Y.: 2009, Measurement and Analysis of an Online Content Voting Network: A Case Study of Digg, *arXiv.org* .
URL: *http://arxiv.org/abs/0909.2706v1*